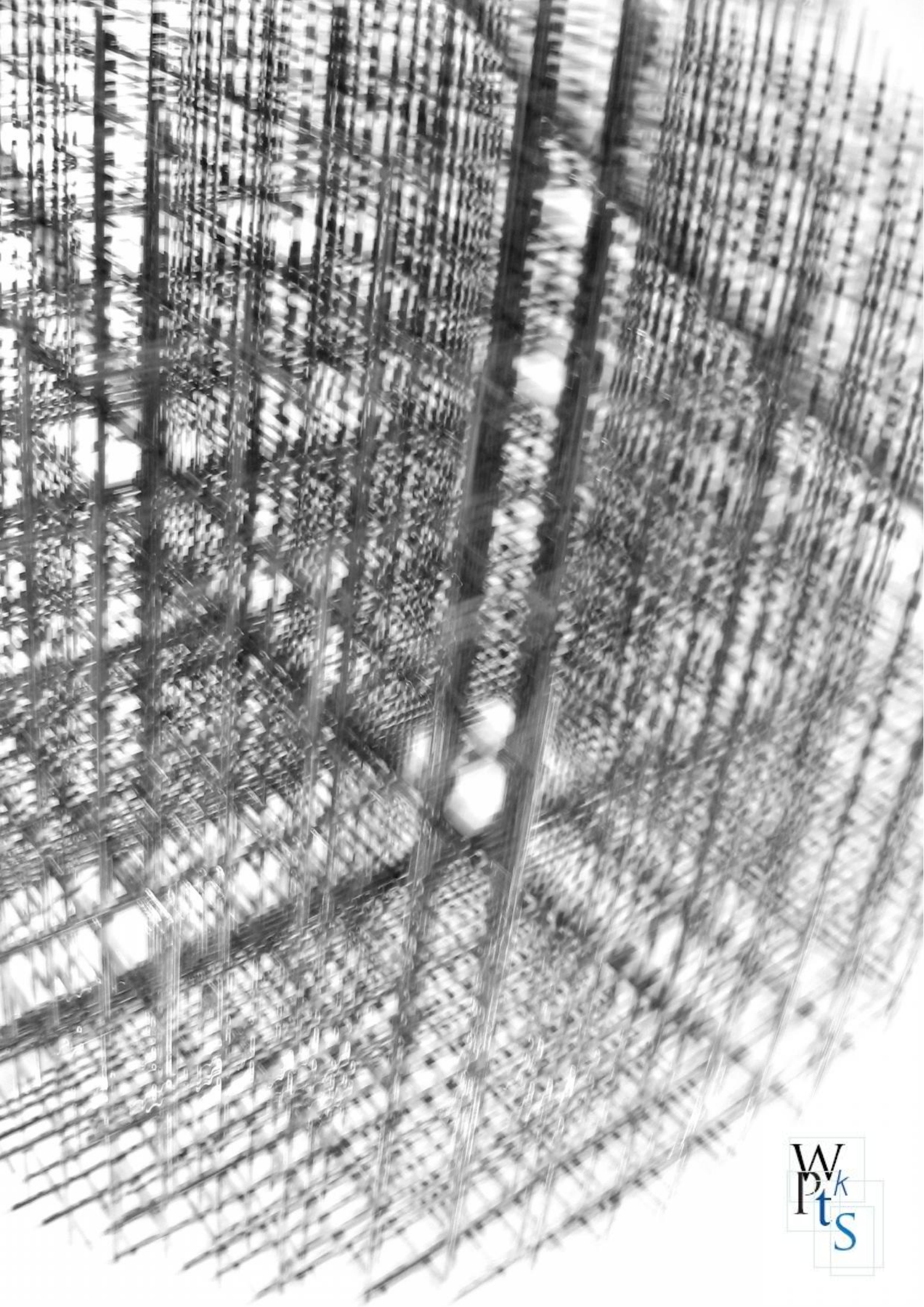
Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung: theoretische Zugänge und soziologisches Anschlusspotential

Marquardt, Manuela

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper



Working Papers
kultur- und techniksoziologische Studien

Volume 10 (1)
no 01/2017

Herausgeber:
Diego Compagna, Stefan Derpmann und Manuela Marquardt
Layout:
Vera Keysers

Kontakt:
diego.compagna@gmail.com
stefan.derpmann@gmail.com
manuela.marquardt@gmx.de

Ein Verzeichnis aller Beiträge befindet sich hier:
www.uni-due.de/wpkts

ISSN 1866-3877
(Working Papers kultur- und techniksoziologische Studien)

Vorwort

Eine soziologische Betrachtung von Technik zeichnet sich unter anderem dadurch aus, dass das Bedingungsverhältnis zwischen den technischen Artefakten und den sozialen Kontexten, in die jene eingebettet sind, als ein interdependentes – zu beiden Seiten hin gleichermaßen konstitutives – angesehen wird. Diesem Wesenszug soziologischer Perspektiven auf Technik trägt der Titel dieser Reihe Rechnung, insofern von einer soziokulturellen Einfärbung von Technik sowie – vice versa – eines Abfärbens von technikinhärenten Merkmalen auf das Soziale auszugehen ist. Darüber hinaus schieben sich zwischen den vielfältigen Kontexten der Forschung, Entwicklung, Herstellung, Gewährleistung und Nutzung zusätzliche Unschärfen ein, die den unterschiedlichen Schwerpunktsetzungen und Orientierungen dieser Kontexte geschuldet sind: In einer hochgradig ausdifferenzierten Gesellschaft ist das Verhältnis von Sozialem und Technik durch je spezifischen Ent- und Rückbettungsdynamiken gekennzeichnet.

Die Reihe Working Papers kultur- und techniksoziologische Studien (WPktS) bietet eine Plattform für den niederschwelligen Austausch mit Kolleg_innen und steht Wissenschaftler_innen und Student_innen aller Universitäten, Fachrichtungen und Institute für die Veröffentlichung ihrer Forschungs- und Qualifikationsarbeiten offen. Der thematische Rahmen ist hierfür mit Absicht breit gewählt und kann mit verschiedensten Darstellungsformen – vom Essay über die Forschungsskizze bis zum Aufsatz – bearbeitet werden.

Die Reihe WPktS erscheint seit 2008; jede Ausgabe kann Online (www.uni-due.de/wpkts) als PDF-Dokument abgerufen werden.

Die Herausgeber
Berlin und Essen, im April 2015

Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung – Theoretische Zugänge und soziologisches Anschlusspotential

Manuela Marquardt

Soziologie (B.A.) / manuelamarquardt@gmx.de

Keywords

Anthropomorphisierung, Roboter, Human-Robot Interaction (HRI), soziale Interaktion

Abstract

Anthropomorphisierung bedeutet die Zuschreibung von menschlichen Eigenschaften hinsichtlich der Gestalt oder des Verhaltens auf nicht-menschliche Entitäten, wie bspw. Götter, Pflanzen, Tiere, Computer oder Roboter. In der Mensch-Roboter Interaktionsforschung ist das Phänomen von besonderer Relevanz, da Roboter aufgrund ihrer Verkörperung und Belebtheit, Anthropomorphisierungen in hohem Maße evozieren. Der Artikel widmet sich theoretischen Zugängen zum Phänomen, integriert diese und zeigt ihre soziologische Anschlussfähigkeit auf. En detail besprochen werden der intentional Stance von Dennett, die Drei-Faktoren-Theorie von Epley et al., ein dynamisches Anthropomorphismusmodell von Lemaignan et al. und ein Ansatz von Persson et al, der Anthropomorphisierung als Mehrebenenphänomen fasst.

1 Einleitung

Anthropomorphisierungen sind ganz alltägliche Phänomene: Von den kleinen Sünden, die der liebe Gott sofort bestraft; über das Haustier, das als vollwertiges Familienmitglied betrachtet wird oder die anspruchsvolle Topfpflanze, die bei mangelnder Zuwendung eingeschnappt ist; bis zum Computer, der einen absichtlich ärgert oder dem Auto, das gut behandelt werden will. Und beim Konzept des sozialen Roboters nehmen sie eine eigenartige Plastizität an. Roboter eignen sich par excellence als Objekte der Anthropomorphisierung und anthropomorphe Zuschreibungen suggerieren darüber hinaus eine Agency (Young et al. 2011) oder etwas, das einer sozialen Beziehung zum Anthropomorphisierungsobjekt nahe kommt. Dabei scheint seit der Moderne – ganz im Gegensatz zu animistischen Kulturen – Einigkeit darüber zu herrschen, dass Menschen als einzig legitime soziale Akteure gelten (Baecker 2011; Lindemann 1999).

Doch was bedeuten diese Zuschreibungen dann? Dienen sie lediglich als verbale Heuristik, um alltagsweltliche Beobachtungen eines gewissen Komplexitätsgrades zu rationalisieren

und intersubjektiv kommunizierbar zu machen? Oder können sie unter gewissen Umständen auch tatsächlich handlungswirksam (bzw. sozial wirksam) werden?

In einer sozialkonstruktivistischen Lesart des Phänomens der Anthropomorphisierung (speziell von Robotern) stellt sich nämlich die Frage, unter welchen Umständen Zuschreibungen menschlicher Eigenschaften auf nicht-menschliche Entitäten auch handlungswirksam werden, denn dann sind diese von besonderem soziologischen Interesse, wenn sie sozialer Natur sind – oder frei nach dem Thomas-Theorem – if men define robots as humanlike (=anthropomorphize robots), they behave socially towards them.

Die Soziologin Sherry Turkle entwickelt den Begriff des evokativen Objektes, um auf die emotionalen und gedanklichen Beziehungen zu verweisen, die Menschen mit Objekten eingehen – diese Objekte sind Gefährten des Gefühlslebens oder sinnstiftende Irritationen (provocation to thought) (Turkle 2007) und insbesondere avancierte Technik wie Computer oder Roboter scheinen sich als evokative Objekte zu eignen (Echterhoff et al. 2006).

Innerhalb der Robotikforschung ist das Phänomen der Anthropomorphisierung von besonderem Interesse, da Roboter allein durch ihre Verkörperung – ganz gleich ob humanoider Gestalt oder nicht – die Tendenz, sie zu anthropomorphisieren, in besonderem Maße hervorrufen. Die Anthropomorphisierung von Robotern kann gleichermaßen als Voraussetzung für die Konstruktion sozialer Roboter gelten. Während es in der Sozialen Robotikforschung bereits eine intensive Diskussion darüber gibt, was das Soziale an Robotern ausmacht und wie der Roboter konstruiert sein muss, um dem gerecht zu werden (Dautenhahn 2007; Meister 2013)[1], soll sich in diesem Artikel dem Phänomen der Anthropomorphisierung – geknüpft an die Annahme, dass diese auch sozialer Natur und damit von besonderem soziologischen Interesse sein kann – zunächst viel basaler und aus einer reinen Zuschreibungsperspektive genähert werden.

1 Dautenhahn referiert verschiedene Definitionen des sozialen Roboters („socially evocative“, „socially situated“, „sociable“, „socially intelligent“ und „socially interactive“) und bündelt diese anhand der drei Pole „robot-centered“, „human-centered“ und „robot cognition centered“ (Dautenhahn 2007: 684f); Meister schlägt eine alternative Klassifizierung anhand der Achsen *robot vs. human centered view* und *flat or no vs. deep modelling of cognition* vor (Meister 2014: 114f).

Denn das rege Interesse an Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung geht gleichzeitig mit einem recht unreflektierten und undifferenzierten Umgang mit dem Konzept einher und es drängt sich der Gedanke auf, dass unter der Sammelbezeichnung Anthropomorphisierung qualitativ doch recht unterschiedliche Aspekte auftauchen. Ziel ist es daher, die verschiedenen Verwendungsweisen des Konzeptes zu sortieren und theoretisch zu verfeinern, um den Weg für eine multidisziplinär (v.a. auch soziologisch) anschlussfähige Theorie der Anthropomorphisierung zu ebnen.

Hierfür wird sich dem Phänomen der Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung anhand folgender drei Fragestellungen genähert:

- Wie wird das Konzept der Anthropomorphisierung theoretisch gefasst?
- Lassen sich ein gemeinsamer Kern rekonstruieren und verschiedene Perspektiven integrieren?
- Welches Anschlusspotential bietet sich für eine soziologische bzw. sozialtheoretische Perspektive?

Im Hauptteil erfolgt zunächst ein knapp einführender, breiter Literaturüberblick zur Anthropomorphisierungsforschung und bisherigen Befunden, speziell im Bereich der Mensch-Roboter Interaktion (2.). Anschließend wird der Forschungsüberblick anhand theoretischer Arbeiten vertieft (3.). Dieser Teil wird eröffnet mit einem in der Anthropomorphismusforschung häufig rezipierten, philosophischen Ansatz von Daniel Dennett zum intentional stance (3.1). Es folgen Arbeiten von Autoren, die sich theoretisch explizit mit dem Phänomen der Anthropomorphisierung beschäftigt haben und hierzu Modelle formulieren: Nicholas Epley‘s, Adam Waytz‘ und John Cacioppo’s Drei-Faktoren-Theorie des Anthropomorphismus (3.2), Per Persson‘s, Jarmo Laaksolahti‘s und Peter Lönnqvist’s Mehrebenenmodell (3.3) und Séverin Lemaignan‘s, Julia Fink‘s, Pierre Dillenbourg’s und Claire Braboszcz‘ dynamisches und kognitives-Phasen-Anthropomorphismusmodell (3.4). Die vorgestellten Ansätze werden im folgenden Kapitel auf ihren gemeinsamen Nenner untersucht und integriert (4.), woraufhin mögliche soziologische bzw. sozialtheoretische Anschlüsse aufgezeigt und das Potential einer genuin soziologischen Perspektive auf Anthropomorphisierung ausgelotet werden sollen (5.). Das Paper schließt mit einem zusammenfassenden Fazit (6.).

2 Literaturüberblick

Das Phänomen Anthropomorphismus kursiert in verschiedenen wissenschaftlichen Debatten und weist Schnittpunkte zu klassischen Disziplinen und Subdisziplinen, aber auch neueren interdisziplinären Forschungsfeldern auf (um nur einige zu nennen: Philosophie, (Sozial-)Psychologie, Ethnologie, Archäologie, Biologie, Kognitionswissenschaften, Künstliche Intelligenz, HCI, Soziale Robotik, HRI). Auch im Marketing beschäftigt man sich intensiv mit dem Thema, da Anthropomorphisierungen bei Marken oder Produkten absatzförderlich sind und die Kundenbindung stärken können (vgl. etwa Guido/Peluso 2015; Hellen/Sääksjärvi 2013).

Anthropomorphe Zuschreibungen sind niederschwellig und können sich auf unterschiedliche Objekte beziehen. Am häufigsten werden Anthropomorphisierungen in der Forschung bei göttlichen/ transzendentalen Entitäten, der Natur im Allgemeinen oder Naturereignissen, Tieren (v.a. Haustieren)[2] und Objekten (v.a. komplexeren technischen Objekten wie Computern oder Robotern) verhandelt. Erklärungen für das Phänomen der Anthropomorphisierung referieren meist auf seine sinnstiftende Funktion. Durch die Übertragung menschenbezogenen (oder sozialen)[3] Wissens wird Unsicherheit reduziert und komplexes Verhaltens erklärbar. Gleichzeitig sind sie jedoch alles andere als beliebig und können von bestimmten Merkmalen des Objektes befördert werden. So gibt es etwa Befunde dazu, dass Bewegtheit ohne sichtbare äußere Einflüsse (vgl. etwa das klassische Experiment von Heider und Simmel (1944) oder Blythe et al. (1999)), die Bewegungsgeschwindigkeit (Morewedge et al.

2 Im Zusammenhang mit der Anthropomorphisierung von Tieren kursieren teilweise skurrile Ansätze mit evolutionärer Perspektive, welche in der Entstehung von Anthropomorphisierungen eine Besonderheit des homo sapiens mit enormem Überlebenswert sehen; gleichzeitig aber auch evolutionäre Selektionsprozesse für die Entstehung von Haustieren ausmachen, welche die Tendenz, sie zu anthropomorphisieren, in besonderem Maße hervorrufen (Serpell 2002). Der Autor bezieht sich auf den Archäologen Mithen, der Anthropomorphismus als definierendes Charakteristikum des homo sapiens ansieht und seine Entstehung im Zusammenhang mit der Entwicklung des reflexiven Bewusstseins ins Mittel- bis Jungpaläolithikum (Altsteinzeit) datiert (40.000 Jahre). Laut Mithen hat die Anthropomorphisierungsneigung enormen Überlebenswert, da der homo sapiens durch die Zuschreibung menschlicher Eigenschaften (Gedanken, Gefühle, Motive etc.) an Tiere einerseits deutlich komplexere Jagdstrategien entwickeln konnte, andererseits aber auch das Zusammenleben mit Tieren als Haus- oder Nutztiere begünstigte (Mithen 1996).

3 Dautenhahn (2004) verhandelt diesen Aspekt in Anknüpfung an ethnologische Forschungen zur „Social Intelligence Hypothesis“, die besagt, dass die menschliche Intelligenz in erster Linie sozialen Ursprungs ist bzw. doch zumindest stark in sozialer Intelligenz verwurzelt ist.

2007) oder die bloße Existenz einer Stimme (Persson et al. 2000) anthropomorphe Zuschreibungen befördern. Darüber hinaus wirken äußerliche Gestaltmerkmale, hauptsächlich gesichts- und körperähnliche Formen anthropomorphisierungsförderlich. Gewisse Objekte evozieren anthropomorphe Zuschreibungen in gewissen Situationen und unter gewissen kulturellen Gegebenheiten bei gewissen Individuen stärker als andere (Epley et al. 2007) und bestimmte Eigenschaften eines Artefaktes können speziell soziale Reaktionen hervorrufen (Dautenhahn 2007).

In der Sozialen Robotik bzw. MRI-Forschung gilt der Anthropomorphisierung eine besondere Aufmerksamkeit, die daher rührt, dass ein besseres Verständnis dieses Phänomens in die Konstruktion und das Design interaktionsfähiger Roboter mit einfließen könnte. Einige Forscher sehen das Potential des Anthropomorphismus als eine Art Sprache für die Mensch-Roboter Interaktion – und zwar durch die konsequente Nutzung der aus anthropomorphen Zuschreibungen resultierenden Erwartungen und deren Übertragung in entsprechende Verhaltensweisen bei sozialen Robotern (Duffy 2003: 181). Derartige Argrumente gehen in Richtung des Leitbildes der intuitiven Mensch-Roboter Interaktion. Intuitiv bedeutet, dass die Interaktion an vorhandenes Wissen des menschlichen Interaktionspartners anknüpft – im Falle der konsequenten Nutzung anthropomorpher Zuschreibungen also an menschenbezogenes Wissen und Wissen über soziale Interaktionen zwischen Menschen. Problematisch ist, dass an dieser Stelle der Zuschreibungsprozess der Anthropomorphisierung (der bei einem menschlichen Beobachter abläuft) häufig mit der Implementierung humanoider Charakteristika [4] (physischer menschenähnlicher Attribute) am Roboter verschwimmt, die Anthropomorphisierungen zwar auslösen können, aber nicht unbedingt müssen bzw. häufig zwar im ersten Moment bestimmte Erwartungen hervorrufen, die dann jedoch schnell enttäuscht werden können.

Der Königsweg in der HRI-Forschung ist daher nicht, Roboter so humanoid wie möglich zu bauen, sondern eine optimale Passung des (physischen und Interaktions-)Designs für die

4 Eine Vielzahl von Studien innerhalb der HRI-Forschung benutzt den Begriff Anthropomorphismus oder anthropomorph jedoch lediglich, um auf gestalterische Merkmale des Roboters aufmerksam zu machen (vgl. etwa Hegel et al. 2008; Riek et al. 2009). Die Gestalt spielt für Anthropomorphisierungen jedoch nicht die wichtigste Rolle. Zahlreiche Ansätze verweisen mittlerweile auf Anthropomorphisierungen ohne menschenähnliche Gestalt des Roboters (vgl. etwa Lemaignan et al. 2014; Złotowski 2015).

Aufgabe und den Anwendungskontext des Roboters zu finden. Die Passungshypothese (matching hypothesis, vgl. Goetz et al. 2003) besagt, dass „die Mensch-Roboter-Interaktion um so erfolgreicher verläuft, je besser das äußere Erscheinungsbild und das Verhalten eines sozialen Roboters mit der Erwartung des Users bzw. der Rolle und Aufgabe des Roboters übereinstimmen“ (Echterhoff et al. 2006: 14).

Ein anderer (weniger HRI-naher) Forschungsstrang beschäftigt sich mit individuellen Differenzen in den Anthropomorphisierungstendenzen und deren Konsequenzen. Die Autoren haben hierfür ein Fragebogeninstrumentarium, den IDAQ (individual differences in anthropomorphism questionnaire), entworfen und Konsequenzen hinsichtlich der Zu- oder Abschreibung menschenähnlicher Eigenschaften untersucht. Ihre Forschungen ergeben im Einklang mit früheren Studien Implikationen anthropomorpher Zuschreibungen (i.S.d. Wahrnehmung eines Agenten als mit Geist/ Verstand ausgestattet) für die Bereiche (1.) Moral[5]; (2.) Verantwortung[6] und (3.) normativ-sozialer Einfluss[7] (vgl. Waytz et al. 2010). Im diesem Zusammenhang tauchen auch häufig (überwiegend psychologische) Forschungen zum umgekehrten Prozess der Dehumanisierung auf (vgl. ausführlich dazu die Dissertationen von Schiffhauer 2015 und Złotowski 2015).

Ausgehend von den Forschungen zur Dehumanisierung (Haslam 2006) werden für die Anthropomorphisierung zunehmend zwei Komplexe dessen, was da zugeschrieben wird, differenziert. Hierbei handelt es sich um Eigenschaften aus dem Bereich der menschlichen Natur (HN, human nature, typically human) und einzigartige menschliche Eigenschaften (HU, human uniqueness, uniquely human) (Ruijten et al. 2014; Schiffhauer 2015; Złotowski 2015)[8].

5 „[…] perceiving an agent to have a mind means that agent is capable of conscious experience and should therefore be treated as a moral agent worthy of care and concern“ (Waytz et al. 2010: 222).

6 „[…] perceiving an agent to have a mind means that the agent is capable of intentional action and can therefore be held responsible for its actions“ (Waytz et al. 2010: 222).

7 „[…] perceiving an agent to have a mind means that the agent is capable of observing, evaluating, and judging a perceiver, thereby serving as a souce of normative social influence on the perceiver“ (Waytz et al. 2010: 222).

8 „There are several aspects that differentiate these two senses of humanness: 1. HU characteristics reflect socialization and culture, while HN characteristics link humans to their inborn biological dispositions. 2. HU characteristics reflect social learning and can vary across cultures and populations. HN is prevalent within populations and universal across different cultures. 3. HN is essential, inherent and natural, while HU may not be perceived as essential. 4. HU involves refinement, civility, morality, and higher cognition. HN involves cognitive flexibility, emotionality, vital agency, and warmth.“ (Złotowski 2015: 28)

Der HU Komplex[9] umfasst höherentwickelte, einzigartige menschliche Eigenschaften, wie etwa Intelligenz, Intentionen, sekundäre Emotionen[10], Selbstkontrolle, Erinnerungs- oder Kommunikationsfähigkeiten, bei deren Absprache im Prozess der Dehumanisierung ein Mensch auf eine Ebene mit einem Tier gestellt wird. Der HN Komplex[11] umfasst primäre Emotionen[12], Geselligkeit und Herzlichkeit, deren Absprache den Menschen laut der diesbezüglichen Forschungen auf eine Ebene mit Objekten oder Automaten stellen würde (Haslam 2006; Złotowski 2015).

Zudem gibt es Hinweise dafür, dass Anthropomorphisierungen von religiösen Entitäten oder Tieren bereitwilliger berichtet werden als Anthropomorphisierungen von Objekten (Chin et al. 2004; Chin et al. 2005)[13]. Auch die Studien aus dem CASA-Bereich (computers are social actors) um Reeves, Nass, Moon und Kollegen weisen immer wieder darauf hin, dass Menschen in der Interaktion mit Computern zwar soziale Heuristiken (wie etwa Höflichkeits- und Reziprozitätsnormen oder Stereotypen) nutzen, sich bei expliziter Nachbefragung jedoch von jeglichem sozialen Verhalten gegenüber einem Computer distanzieren (vgl. etwa Reeves/Nass 1998; Nass/Moon 2000). Ähnliche Befunde gibt es für die MRI-Forschung. So stellten Fussel et al. (2008) etwa fest, dass die Anthropomorphisierung von Robotern auf verschiedenen Abstraktionsniveaus unterschiedlich stark ausfällt: Während in freien Beschreibungen einer MRI-Situation und anschließenden, schnell zu bearbeitenden ja/nein-Adjektivlisten mit menschlichen und nicht-menschlichen Eigenschaften der Roboter in beträchtlichem Maße anthropomorphisiert wurde, haben die Versuchspersonen abstrakteren

9 Haslam fasst hierunter die Bereiche „civility“, „refinement“, „moral sensibility“, „rationality, logic“ und „maturity“ (Haslam 2006: 257).

10 Sekundäre Emotionen werden üblicherweise als abgeleitete, komplexere oder sozialisierte Emotionen beschrieben, die aus einem Grundgefühl hervorgehen können. Sie umfassen bspw. Scham, Kummer, Verlegenheit, Eifersucht, Verachtung, Lust oder Schuld.

11 Haslam fasst hierunter „emotional responsiveness“, „interpersonal warmth“, „cognitive openness“, „agency, individuality“ und „depth“ (Haslam 2006: 257).

12 Zu den primären Emotionen werden üblicherweise Freude, Zorn, Furcht, Ekel, Traurigkeit und Überraschung gezählt, sie treten früh in der kindlichen Entwicklung und über alle Kulturen hinweg auf.

13 Die Autoren entwickelten einen 208-Item-Fragebogen (ATS – anthropomorphic tendencies scale) mit 20 Fragen (nach Einstellungen, Intentionen, Verhaltensweisen, Emotionalität und Zuschreibungen) zu je 12 Objekten (Computer, Stofftier, Mikrowelle, Glück, Zimmerpflanze, Ozean, Gott bzw. höhere Macht, Auto, Rucksack, Insekt, Haustier, Magen) und testeten diesen an 1462 Studenten. Sie berichten faktorenanalytisch vier unabhängige Arten von Anthropomorphisierungstendenzen, nämlich (1.) extreme Anthropomorphisierungstendenzen, (2.) Anthropomorphismus göttlicher Entitäten, (3.) Anthropomorphismus von Haustieren und (4.) eher emotional aufgeladene Reaktionen auf diverse nicht-menschliche Entitäten.

Aussagen über Stimmungen, Absichten oder Gefühle eines Roboters deutlich weniger zugestimmt (Fussell et al. 2008). Das legt die Vermutung nahe, dass Anthropomorphisierungen auf unterschiedlichen Ebenen wirksam werden können bzw. dass es einen Unterschied zwischen impliziten und expliziten Anthropomorphisierungen gibt, die Złotowski mit unterschiedlichen Modi der Informationsverarbeitung (Type I und Type II Processing) in Verbindung bringt (Złotowski 2015)[14].

Diese Befunde zeigen den Bedarf an begrifflich-theoretischen Schärfungen (Epley et al. 2007; Persson et al. 2000; Lemaignan et al. 2014) und knüpfen an einen beginnenden Methodendiskurs in der Anthropomorphismusforschung an, der nach der Adäquanz verschiedener Messverfahren für anthropomorphe Zuschreibungen fragt (Kamide et al. 2013; Ruijten et al. 2014; Weiss/Bartneck 2015; Złotowski 2015).

3 Theoretische Arbeiten zur Anthropomorphisierung

Nach dieser breiten Übersicht über Forschungsschwerpunkte und Befunde der Anthropomorphismusforschung, werden nun vier theoretische Arbeiten zum Phänomen der Anthropomorphisierung vorgestellt. Üblicherweise herrscht in den empirischen Studien, die sich mit Anthropomorphisierung von Robotern beschäftigen, eine Theoriearmut und die Autoren gehen doch recht undifferenziert mit dem Konzept (und seiner Messung!) um. Bei genauerer Recherche lassen sich aber doch eine Hand voll Arbeiten finden, welche ganz unterschiedliche Aspekte des Phänomens theoretisch fassen. Eröffnet wird die Bühne mit Daniel Dennett, einem Philosophen des Geistes, dessen intentional systems theory v.a. in den Kognitionswissenschaften rezipiert wurde und der auch in MRI-Arbeiten immer wieder als theore-

14 Seine Studie kommt zu dem Schluss, dass Anthropomorphisierung ein Type II Prozess sei (also das was Esser (2006) als RC-Modus bezeichnet), da in seinen Daten kein signifikanter Unterschied hinsichtlich der Anthropo-morphisierungsbereitwilligkeit zwischen den beiden Konditionen hohe vs. niedrige Motivation für Type II Prozessierung ersichtlich war. Daraus zieht er zudem den Schluss, Selbstbeurteilungsfragebögen seien ein angemessenes Messinstrument für Anthropomorphismus. Seine Stichprobe besteht jedoch aus 40 japanischen Studenten, denen für ihre Teilnahme 2000 Yen (ca. 16 Euro) bezahlt wurde und die Manipulation für die Experimentalkonditionen hohe vs. niedrige Motivation für Type II Prozessierung bestand darin, der einen Hälfte der Probanden zu sagen, man wölle nach der Studie ihre Task Performance und die Fragebogenantworten mit ihnen diskutieren , während die andere Hälfte nur über die Anonymisierung ihrer Daten informiert wurde. Złotowski selbst merkt in den Limitierungen an, dass keine Manipulationskontrolle für die beiden Konditionen stattgefunden hat und dass die Resultate eventuell auch kulturell bedingt oder (in die entgegengesetzte Richtung) sozial erwünscht sein könnten (Złotowski 2015).

tischer Verweis auftaucht; gefolgt von der wohl prominentesten (weil elaboriertesten) Anthropomorphismustheorie von Epley und Kollegen, welche sich mit den (sozial-) psychologischen Determinanten anthropomorpher Zuschreibungen beschäftigen. Es folgen in der MRI-Forschung kaum rezipierte Ansätze einer schwedischen Forschergruppe (Persson et al. 2000), welche Anthropomorphismus als Phänomen beschreiben, das auf verschiedenen (gewissermaßen hierarchischen) Ebenen operiert und einer schweizerischen Forschergruppe (Lemaignan et al. 2014), die sich insbesondere mit Langzeitstudien beschäftigen und Anthropomorphismus als dynamisches Phänomen fassen, das sich abgesehen von temporären Eruptionen mit der Interaktionsdauer stabilisiert.

3.1 Der Intentional Stance

Im Zusammenhang mit Anthropomorphismus wird häufig auf einen Aufsatz aus den 70ern verwiesen, der sich aus philosophischer Perspektive mit verschiedenen Grundhaltungen beschäftigt, die Menschen einnehmen können, um das Verhalten eines Systems zu verstehen und vorhersagbar zu machen. Daniel Dennett ist ein US-amerikanischer Philosoph, der sich intensiv mit der Philosophie des Geistes beschäftigt hat. Seine Theorie intentionaler Systeme dreht sich darum, wie und warum es möglich ist, das Verhalten zahlreicher komplizierter Dinge zu verstehen, indem man sie als Agenten betrachtet (Dennett 2009). Dennett definiert intentionale Systeme als „[…] system whose behavior can be (at least sometimes) explained and predicted by relying on ascriptions to the system of beliefs and desires (and hopes, fears, intentions, hunches,…)“ (Dennett 1971: 87). Das System gilt demnach nur als intentionales System in Relation zu jemandem, der die benannte Strategie, den intentional stance, einnimmt, um das Verhalten des Systems erklärbar zu machen. Die Grundhaltung, die in der HRI-Literatur mit Anthropomorphisierung in Zusammenhang gebracht wird, ist der *intentional stance*, also die Erklärung des Verhaltens eines Systems auf der Basis von Überzeugungen, Wünschen, Intentionen oder anderen bedeutungsvollen Geisteszuständen. Neben dieser differenziert Dennett zwei weitere Grundhaltungen (stances), die Menschen dabei helfen, Verhalten zu verstehen und vorherzusagen: den physical stance und den design stance. Während der physical stance Vorhersagen auf der Basis physikalischer Charakteristika von Systemen vornimmt, zielt der design stance auf Vorhersagen basierend auf Design- und Funktionalitätsaspekten eines Systems (Dennett 2009;

1971).[15]

Diese unterschiedlichen Grundhaltungen oder Einstellungen umschreiben im Prinzip, welche Art von Wissen für einen induktiven Schluss herangezogen wird, um beobachtbares Verhalten zu generalisieren und damit prognostizierbar zu machen, um entsprechend darauf reagieren zu können.[16] Während es sich beim physical stance um Wissen aus dem naturwissenschaftlichen Bereich handelt (also bspw. Massen, Gravitation, Beschleunigung), genügt beim design stance Wissen über die Funktion einer Entität (gekoppelt mit der Annahme, dass kein Defekt vorliegt); beim intentional stance ist es das Wissen einer *Theory of Mind*, also das Wissen über (eigene) Geisteszustände und deren Übertragung auf andere Entitäten. Dabei behauptet Dennett, für die Intentional Systems Theory mache es keinen Unterschied, ob es sich dabei um *echte* (original, literal, intrinsic) oder lediglich zugeschriebene (derived, metaphorical, as if) Intentionen handle (Dennett 2009: 7f).

Damit beinhaltet der intentional stance u.a. das, was gemeinhin als Anthropomorphisierung gilt, nämlich die Zuschreibung menschlicher Eigenschaften auf nicht-menschliche Entitäten, sofern man menschliche Eigenschaften mit gewissen Geisteszuständen gleichsetzt. Nach Dennett ist der intentional stance jedoch nicht auf nicht-menschliche Entitäten beschränkt, da jeder Mensch nach seiner Theorie auch ein intentionales System darstellt. Der große Vorteil in der Einnahme des intentional stance liegt in der Rationalisierbarkeit und dadurch gewonnenen Vorhersagekraft beobachtbaren Verhaltens. Gleichzeitig stellt sich jedoch die Frage, ob die Grundhaltung des intentional stance das Phänomen der Anthropomorphisierung ausreichend fassen kann oder ob sie es nicht vielmehr lediglich funktional erklärt. Die folgenden Ansätze widmen sich ganz explizit der Anthropomorphisierung und beschäftigen sich damit, welche Faktoren anthropomorphe Zuschreibungen befördern, was genau das menschliche oder menschenähnliche ist, das da zugeschrieben wird und wie sich Anthropomorphisierungen eines Roboters über die Interaktionsdauer hinweg stabilisieren.

15 Dieser gestufte Zuschreibungsbegriff weist gewisse Analogien zum gestuften Handlungsbegriff von Rammert und Schulz-Schaeffer auf: Der physical stance kommt auf der Stufe der verändernden Wirksamkeit zum Einsatz, der design stance auf der Stufe des auch-anders-handeln-könnens und der intentional stance auf der Stufe intentionaler Erklärungen (Rammert/Schulz-Schaeffer 2002).

16 Es handelt sich genauer gesagt um qualitative Induktionen, bei denen von bekannten Eigenschaften eines Ereignisses auf weitere, nicht-beobachtbare geschlossen wird, indem der Einzelfall unter eine bekannte Regel subsumiert wird, die auf Fälle mit ähnlichen Eigenschaften angewandt wird (Reichertz 2013: 18f).

3.2 Drei-Faktoren-Theorie des Anthropomorphismus

Nicholas Epley, Adam Wayth und John Cacioppo von der Chicago Universität formulieren unter Rekurs auf einen reichen Fundus an psychologischen Forschungen eine Drei-Faktoren-Theorie des Anthropomorphismus, in der sie verschiedene Dispositions-, Situations-, Entwicklungs- und Kulturfaktoren als mögliche Einflussvariablen auf die Faktoren „elicited agent knowledge", „effectance motivation" und „sociality motivation" vorschlagen.[17] Mit ihrer Theorie setzen sie sich zum Ziel, die Variabilität der Anthropomorphisierungstendenzen systematisch zu erklären und vorherzusagen. Dabei definieren sie Anthropomorphismus folgendermaßen: „Anthropomorphism describes the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike charakteristics, motivations, intentions, or emotions."[18] (Epley et al. 2007: 864)

Als zentrale – wenngleich nicht erschöpfende – menschenähnliche Zuschreibungskategorien nennen sie geistesbezogene Vorstellungen (mind) wie bewusste Erfahrung, Metakognition und Intentionen. Darüber hinaus gehen sie auch von der Zuschreibung emotionaler Zustände, verhaltensbezogener Charakteristika oder menschenähnlicher Formen aus. Die Zuschreibung ist dabei in erster Linie eine Inferenz auf die nicht-beobachtbaren Charakteristika der nicht-menschlichen Entität, keine bloße Verhaltensbeschreibung.

Anthropomorphisierung funktioniert psychologisch analog zu anderen Prozessen der induktiven Inferenz über basale kognitive Operationen des Wissenserwerbs, der Aktivierung gespeicherten Wissens und der Anwendung des aktivierten Wissens auf ein bestimmtes Zielobjekt. Bei der Anwendung gilt gemeinhin, dass die am leichtesten abrufbare Information die Richtung der finalen Beurteilung am stärksten beeinflusst; es können jedoch auch Kor-

17 „Some nonhuman agents are anthropomorphized more than others. Children appear to anthropomorphize nonhuman agents more than adults. Some people anthropomorphize nonhuman agents more than other people. Some situations seem to elicit anthropomorphic beliefs more than others, and some cultures seem especially fond of anthropomorphic descriptions compared with others." (Epley et al. 2007: 865)

18 An anderer Stelle verwenden die Autoren auch eine der gängigeren Definitionen für Anthropomorphismus als „attribution of human characteristics or traits to nonhuman agents" (Epley et al. 2007: 865).

rekturprozesse initiiert werden, die versuchen alternative Wissensstrukturen zu integrieren.[19] Für den kognitiven Prozess gibt es laut der Autoren demnach drei zentrale Stellschrauben: die Wahrscheinlichkeit der Aktivierung von Wissen über Menschen, die Wahrscheinlichkeit der Korrektur oder Anpassung anthropomorpher Repräsentationen mit nicht-anthropomorphem Wissen und die Wahrscheinlichkeit der Anwendung der aktivierten und ggf. korrigierten anthropomorphen Repräsentationen auf nicht-menschliche Agenten (Epley et al. 2007: 865).

Die kognitiven Prozesse fassen die Autoren im Faktor „elicitation of agent knowledge", der hier mit „Auslösung menschenbezogenen Wissens" übersetzt werden soll. Hier spielen also die (chronische oder situationale) Zugänglichkeit anthropomorpher Repräsentationen, deren Aktivierung, Korrektur/ Integration und Anwendung eine Rolle. Dieses Wissen ist gemeinhin sehr detailliert und gut zugänglich, es sei denn es liegen alternative Wissensbestände über nicht-menschliche Agenten vor. Der kognitive Prozess kann durch motivationale Einflüsse gelenkt oder verändert werden.[20]

Als zweiten Faktor benennen die Autoren die „effectance motivation". Die Wirksamkeitsmotivation bzw. das Effektanzmotiv bezeichnet die Motivation, effektiv mit der Umwelt zu interagieren, besonders in Situationen, die unterdeterminiert sind. Bezogen auf Anthropomorphismus geht es bei diesem Motiv also darum, effektiv mit einem nicht-menschlichen Agenten zu interagieren, d.h. gegenwärtige (komplexe) Stimuli zu erklären und zukünftiges Verhalten vorhersagbar zu machen. In diesem Zusammenhang beziehen sich die Autoren auch auf Dennetts Ansatz und seine utilitaristische Funktion, da er die Zweckdienlichkeit von Anthropomorphisierungen (unabhängig von der konkreten Beschaffenheit der nicht-menschlichen Entitäten) für eine effektive Auseinandersetzung mit der Umwelt hervorhebt (Epley et al. 2007: 871f). Durch die Zuschreibung menschlicher Charakteristika und Motiva-

19 „The ready accessibility of self-knowledge and one's own phenomenology makes an anthropomorphic inference a likely intuitive anchor or starting point when reasoning about nonhuman agents, and correction of this anchor is possible to the extent that people are motivated and able to do so" (Epley et al. 2007: 869).

20 Die Autoren begründen ihre Differenzierung nach kognitiven und motivationalen Einflüssen mit den unterschiedlichen Wirkmechanismen: Während motivationale Einflüsse wie Triebe bei Deprivation verstärkt werden und nachlassen, sobald der Antrieb befriedigt ist, werden kognitive Einflüsse mit ihrer Aktivierung verstärkt und lassen dann mit der Zeit nach (Epley et al. 2007: 871).

tionen wird die Sinnhaftigkeit der Handlungen des nicht-menschlichen Agenten erhöht, Unsicherheit reduziert und Vertrauen in die Vorhersagbarkeit des Agenten gewonnen. Mögliche Einflussfaktoren auf die Anthropomorphisierungstendenz sind hier bspw. die Angst vor Unsicherheit und die Wichtigkeit, das Verhalten eines Agenten in einer konkreten Situation vorherzusagen.

Der dritte Faktor ist die „sociality motivation“, welche das grundlegende menschliche Bedürfnis und den Wunsch nach sozialer Zugehörigkeit umfasst. Durch Anthropomorphismus können menschenähnliche Bindungen zu nicht-menschlichen Agenten aufgebaut werden, welche dieses Bedürfnis befriedigen. Ein möglicher motivationaler Einflussfaktor auf die Anthropomorphisierungstendenz ist demnach ein Mangel an sozialen Bindungen zu anderen Menschen. Beide motivationalen Faktoren erhöhen bei starker Ausprägung jeweils die Zugänglichkeit, vermindern die Korrektur und steigern die Anwendung anthropomorphen Wissens auf nicht-menschliche Agenten (Epley et al. 2007).

Tabelle 1 zeigt mögliche Einflussvariablen auf diese drei Faktoren, sortiert nach den Kategorien dispositional, situational, entwicklungsbezogen und kulturell. Bei Epley et al. wird jede Variable ausführlich dargestellt und mit entsprechender psychologischer Forschung untermauert. Soziologisch von besonderem Interesse sind die situationalen und kulturellen Variablen. Situationale Einflussvariablen auf die Anthropomorphisierungstendenz wären demnach etwa die wahrgenommene Ähnlichkeit der nicht-menschlichen Entität (v.a. in Bezug auf die Bewegung/-sgeschwindigkeit und die Morphologie, Epley et al. 2007: 869), die Erwartung an die Interaktion bzw. deren Prognostizierbarkeit (situationale Unsicherheit, bspw. durch Erwartungsverletzungen, kann durch die (Re-)Formulierung anthropomorpher Erwartungen reduziert werden und das Verhalten eines nicht-menschlichen Agenten prognostizieren, Epley et al. 2007: 872f) und die situative soziale Trennung bzw. Abgeschiedenheit (als Beispiel nennen die Autoren u.a. Tom Hanks im Film „Cast Away“, der einen Volleyball namens Wilson anthropomorphisiert, Epley et al. 2007: 876f).

Kulturell variieren Anthropomorphisierungstendenzen je nach Erfahrungen, Normen und Ideologien (indem diese Erwerb und Zugänglichkeit von Repräsentationen des Selbst, anderer Menschen und nicht-menschlicher Agenten beeinflussen, Epley et al. 2007: 870f), den kulturellen Differenzen der Unsicherheitsvermeidung (Epley et al. 2007: 874f) und einem

eher individualistischen oder kollektivistischen Wertesystem (Epley et al. stellen die Hypothese auf, dass in kollektivistischen Kulturen weniger chronische Einsamkeit vorherrscht, diese aber dann auch stärker auf kurzzeitige soziale Isolation oder Exklusion reagieren, vgl. 877).

unabhängige Variablen-kategorien	**Auslösung menschen-bezogenen Wissens**	**Wirksamkeits-motivation**	**Sozialitäts-motivation**
dispositional	„Need for Cognition“[21]	„Need for Closure“[22], „Desire for Control“[23]	chronische Einsamkeit
situational	wahrgenommene Ähnlichkeit	erwartete Interaktion, scheinbare Vorhersagbarkeit	soziale Trennung
entwicklungs-bezogen	Erwerb von alternativen Theorien	erreichte Kompetenz	Bindung (attachment)
kulturell	Erfahrungen, Normen und Ideologien	Vermeidung von Unsicherheit	Individualismus/ Kollektivismus

Tabelle 1: Einflussfaktoren auf die psychologischen Schlüsseldeterminanten der Drei-Faktoren-Theorie (Übersetzung in Anlehnung an Schiffhauer 2015, S. 13)

Auch Epley und Kollegen unterscheiden zwischen schwächeren (eher metaphorischen) und stärkeren Formen des Anthropomorphismus, sind jedoch überzeugt, beide Varianten mit ihrer Theorie erklären zu können. Darüber hinaus weisen sie darauf hin, dass schwächere Anthropomorphisierungen, die zunächst lediglich im Sinne einer verbalen Heuristik fungieren, stärker ins Gewicht fallen, als man intuitiv annehmen würde, denn auch Metaphern können einen Einfluss auf Verhalten ausüben (Epley et al. 2007: 867). „The difference between weak and strong versions of anthropomorphism, we suggest, is simply a matter of degree regarding the strength and behavioral consequences of a belief, not a fundamental difference in kind.“ (Epley et al. 2007: 867)

21 “Need for Cognition“ bezeichnet die individuelle Tendenz, anstrengendere kognitive Verarbeitungsprozesse anzustellen und an ihnen Gefallen zu finden; Bless et al. übersetzen NFC mit Spaß am Denken (Bless et al. 1994).

22 “Need for Closure“ oder „Need for Cognitive Closure“ bezeichnen das Bedürfnis nach kognitiver Geschlossenheit bzw. umgekehrt die Aversion gegenüber Ambiguität.

23 „Desire for Control“ kann mit Kontrollüberzeugung übersetzt werden und umfasst die Disposition, ob die Kontrolle über das Auftreten von Ereignissen eher innerhalb oder außerhalb des Individuums angesiedelt wird.

Epley und Kollegen bringen mit ihrem Ansatz im Vergleich zu Dennett einen ganz zentralen Punkt mit in die Debatte: das Wissen. Demnach sind Anthropomorphisierungen insbesondere dann von Relevanz, wenn nur unzureichendes Wissen über die nicht-menschliche Entität vorliegt. Zudem kann die Zugänglichkeit von Wissensbeständen durch bestimmte situationale Cues gelenkt werden, was insbesondere für MRI-Laborexperimente von Bedeutung ist. Ihr Hinweis auf das Effektanzmotiv deckt sich mehr oder weniger mit Dennett und fällt v.a. in HRI-Experimenten stark ins Gewicht, da sich diese Situationen im Normalfall durch Neuheit, Unbekanntheit und ggf. Unterdeterminiertheit auszeichnen. Das Bedürfnis nach sozialer Anerkennung ist für die Robotikforschung insbesondere in der Diskussion um robot companions von Interesse, die eine ähnliche Rolle wie Haustiere einnehmen können und Menschen somit in gewissen Maßen als Substitut für menschliche soziale Kontakte dienen können.

Trotz des hohen Elaborationsgrades der Drei-Faktoren-Theorie des Anthropomorphismus differenzieren Epley et al. nicht weiter, *was* es ist, das da zugeschrieben wird und wie sich die Zuschreibungen dynamisch verändern (obgleich mit ihrer Theorie die Anthropomorphisierungsdynamik, wie später zu zeigen sein wird, gut erklärbar ist). Dies wird der inhaltliche Schwerpunkt der folgenden beiden, in der HRI-Forschung kaum rezipierten Ansätze sein.

3.3 Anthropomorphismus als Mehrebenenphänomen

Persson et al. haben bereits im Jahr 2000 ein Paper publiziert, in dem sie ziemlich genau differenzieren, *was* bei Anthropomorphisierungen zugeschrieben wird, denn „[a]*nthropomorphism* means different things on different levels“ (Persson et al. 2000: 2). Daher geben sie auch keine inhaltliche Definition von Anthropomorphismus, sondern lediglich eine funktionale, in dem Sinne, dass Anthropomorphismus zur Sinndeutung komplexen Verhaltens fungiert. Sie schreiben dazu selbst: „Anthropomorphism has a fundamental ’sense-making’ function. Thus, an interface is not anthropomorphic per se, but only in so far as it gives rise to anthropomorphic processes in a given user and situation.“ (Persson et al. 2000: 1) Mit dieser Auffassung des Phänomens lokalisieren sie Anthropomorphismus in der Interaktion zwischen dem Menschen und der nicht-menschlichen Entität. Um nochmals ihre Worte zu verwenden: „Anthropomorphism resides neither wholly in ’objective reality’, nor wholly in the

mind of the observer. It arises in the interaction between a set of anthropomorphic expectations and external reality“ (Persson et al. 2000: 1).

Sie machen also aus psychologischer Sicht verschiedene hierarchisch angeordnete Ebenen aus, auf denen Anthropomorphismus operieren kann. Die erste Ebene bildet die primitive Kategorisierung, auf welcher zunächst Bewegtheit, die visuelle Erscheinung (menschliche Gestalt oder Bewegung) und Stimmen als besonders starke Cues für spontane Anthropomorphisierungen fungieren und welche die Basis für Zuschreibungen auf höheren Ebenen bildet. Die zweite Ebene nennen die Autoren primitive Psychologie. Hier werden sehr basale Erwartungen hinsichtlich der Selbsterhaltung, primärer Bedürfnisse (bspw. Hunger, Durst, Schlaf) oder Triebe (bspw. Sexualtrieb), aber auch der Wahrnehmung und etwa des Empfindens von Schmerz formuliert. Auf der dritten Ebene der Alltagspsychologie (folk-psychology) werden die Erwartungen komplexer, indem Annahmen über die Beziehungen von Wahrnehmungen, Einstellungen, Zielen, Intentionen, Emotionen und Handlungen zur Erklärung von beobachtbarem Verhalten getroffen werden. An vierter Stelle können sogenannte Traits bzw. Wesenszüge zugeschrieben werden. Durch Wesenszüge kann der Eindruck einer Person oder eines Systems kompakt zusammengefasst werden. Sie sind beständiger und stabiler als die Zuschreibung innerer Zustände der Alltagspsychologie, fassen jedoch gleichzeitig komplexe alltagspsychologische Anthropomorphisierungsprozesse in einem Begriff zusammen (bspw. implizieren Beschreibungen einer Entität als schüchtern, aggressiv oder unbeholfen bestimmte Konstellationen innerer Zustände zur Erklärung ihres Verhaltens). Als fünfte Ebene von Anthropomorphismus nennen die Autoren soziale Rollen, welche normative Erwartungen beinhalten und bspw. in bestimmte soziale event schemas oder scripts eingebettet sein können. An dieser Stelle spielt die gesamte Situation eine zentrale Rolle, die Hinweise für die Zuschreibung bestimmter Rollen gibt. Als Beispiel nennen die Autoren zudem soziale Stereotype, bei denen gewisse Wesenszüge einer Gruppe von Menschen zugeschrieben und emotional oder moralisch evaluiert werden. Stereotype beinhalten zudem eine ikonographische Dimension, denn sie sind häufig an auffällige äußere Merkmale geknüpft, welche vor allem bei Erstkontakt wirksam werden.

Die Autoren benennen als letzte Ebene zusätzlich den emotionalen Anthropomorphismus und behaupten, alle anderen Ebenen hätten bislang kognitive Erwartungen thematisiert, die

Menschen auf verhaltensmäßig komplexe Systeme projizieren. Mit dieser letzten Ebene wollen sie daher der emotionalen Seite von Anthropomorphisierungen Rechnung tragen. Dieser Vorschlag erscheint jedoch nicht sonderlich trennscharf, da Emotionen auch auf den anderen Zuschreibungsebenen auftauchen können.[24] Dennoch ist ihr Modell (abgesehen vom emotionalen Anthropomorphismus) für die MRI-Forschung höchstinteressant, da auf diese Weise qualitativen Unterschieden in den Anthropomorphisierungen Rechnung getragen werden kann.[25] Somit kann empirisch nicht nur festgestellt werden, ob eine anthropomorphe Zuschreibung stattfindet, sondern auch auf welcher Ebene sie operiert. Unzureichend äußern sie sich jedoch hinsichtlich der Handlungswirksamkeit anthropomorpher Zuschreibungen, welche Epley et al als „strength and behavioral consequences of a belief" (Epley et al. 2007: 867) umschrieben haben. Diese wird in der soziologischen bzw. sozialtheoretischen Reflexion nochmals aufgegriffen werden.

Um den Literaturüberblick abzuschließen, wird nun als letzter Ansatz ein Anthropomorphismusmodell vorgestellt, das versucht, anthropomorphe Zuschreibungen über die Zeit hinweg zu modellieren und mit den dabei ablaufenden kognitiven Prozessen zu parallelisieren.

3.4 Dynamisches Modell des Anthropomorphismus

Lemaignan et al. (2014) stellen ein kognitionswissenschaftlich inspiriertes Modell des Anthropomorphismus zur Diskussion, in dem sie insbesondere der Dynamik anthropomorpher Zuschreibungen über die Interaktionsdauer hinweg Rechnung tragen wollen. Sie definieren Anthropomorphismus als soziales Phänomen, das in der Interaktion zwischen dem Roboter und dem Nutzer entsteht und grenzen es von anthropomorphem Design (als statisches Set menschenähnlicher Eigenschaften eines Roboters – wie etwa Gestalt, Sprachfähigkeiten,

24 Auch in der Emotionssoziologie wird versucht, den Dualismus zwischen Vernunft und Emotion aufzulösen und deutlich zu machen, dass Emotionen bei jeglichen kognitiven bzw. Wahrnehmungs-Prozessen eine Rolle spielen können. Sogar Hartmut Esser hat den Versuch unternommen, Emotionen in sein RC-orientiertes Modell der Frame Selektion zu integrieren. Diese können sowohl im AS-, als auch im RC-Modus eine Rolle spielen (Esser 2006).

25 Einen ähnlichen Versuch haben Fussell et al. (2008) unternommen, in dem sie anhand eines linguistischen Kategorisierungsmodells verschiedene Ebenen anthropomorpher Zuschreibungen ausgemacht haben. Sie grenzen „descriptive action verbs" (wie bspw. schlagen) von „interpretative action verbs" (wie bspw. misshandeln) von „state verbs" (wie bspw. hassen) und „trait adjectives" (wie bspw. aggressiv) als zunehmend anthropomorphen Zuschreibungen ab (Fussell et al. 2008: 146).

Gesichtsausdruck) ab (Lemaignan et al. 2014: 1).

Ihr dynamisches Anthropomorphismusmodell stellt die verschiedene Phasen der Interaktion über die Zeit hinweg (Initialisierung, Gewöhnung, Stabilisierung) und die kognitiven Prozesse der sukzessiven Bildung eines mentalen Modells des Roboters (Phase I: sub-kognitiver Anthropomorphismus, Phase II: Projektion existenter mentaler Modelle, Phase III: angepasste mentale Modelle) zunächst lose nebeneinander (vgl. **Fehler! Verweisquelle konnte nicht gefunden werden.**).

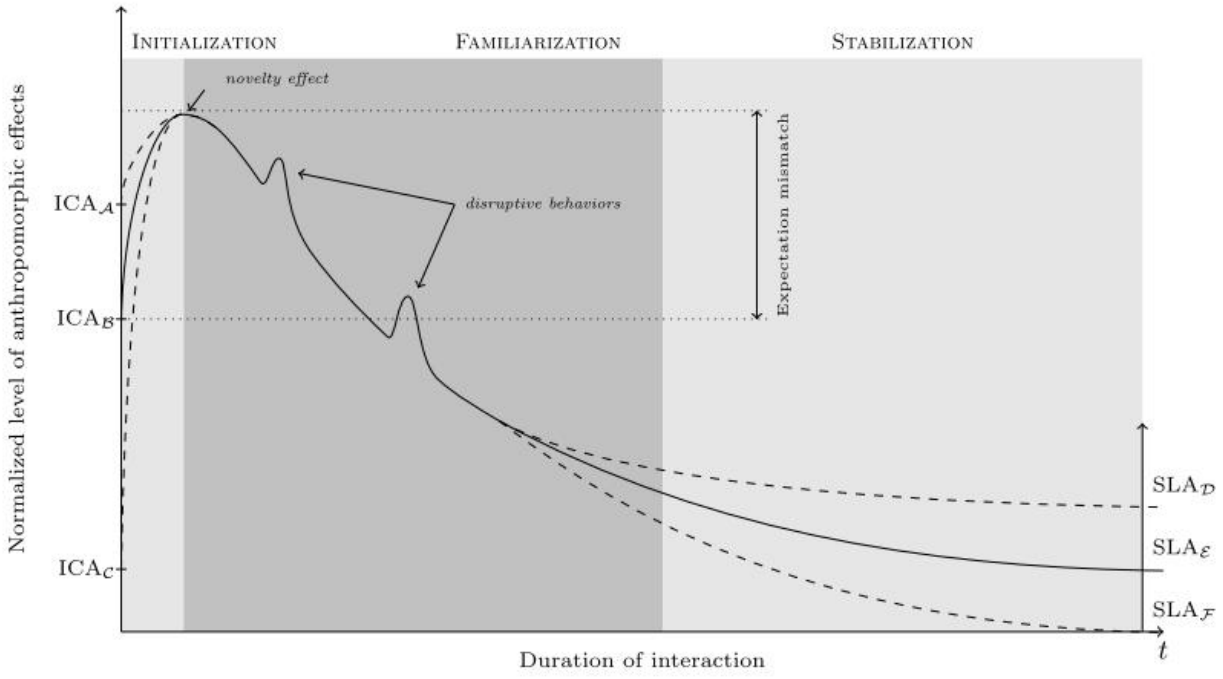


Abbildung 1: Dynamisches Anthropomorphismusmodell (Initialisierung, Gewöhnung, Stabilisierung), aus Lemaignan et al. 2014, S. 2

Zur Beschreibung der Anthropomorphisierungstendenz wählen sie ein – inhaltlich nicht weiter definiertes – Maß, das sie „normalized level of anthropomorphic effects" nennen. Diese anthropomorphen Effekte sind beobachtbare Manifestationen des Anthropomorphisierungsprozesses und bilden sich über die Zeit hinweg in nicht-monotoner Art und Weise aus, wobei am Minimum keinerlei anthropomorphe Effekte beobachtbar sind und der standardisierte Wert am Maximum (dieses Maximum nennen die Autoren den „novelty effect") von der jeweiligen Kombination mensch-zentrierter, roboter-zentrierter und situations-zentrierter Fak-

toren abhängt. Die Form der Kurve lässt sich anhand des ICA (initial capital of anthropomorphism), des novelty effects und des SLA (stabilized level of anthropomorphism) beschreiben. In das ICA fließen drei Faktoren ein, welche laut der Autoren a priori das Anthropomorphisierungspotential festlegen: mensch-zentrierte Faktoren[26] (bspw. Alter, Geschlecht, kultureller oder beruflicher Hintergrund), roboter-zentrierte Faktoren (bspw. Form, Verhalten und Interaktionsmodalitäten) und situations-zentrierte Faktoren (bspw. räumlicher Kontext, Verwendungszweck, Rolle des Roboters).

Die Initialisierungsphase dauert von wenigen Sekunden bis einigen Stunden und zeichnet sich durch einen Anstieg des ICA bis zum Maximum der beobachtbaren anthropomorphen Effekte aus. Daran schließt sich die Gewöhnungsphase an, welche die ersten Tage andauern kann und in der sich durch Beobachtung und Interaktionserfahrungen mit dem Roboter sukzessive ein mentales Modell des Roboters beim menschlichen Interaktionspartner bildet. In dieser Phase lassen die anthropomorphen Effekte nach, da sich der Nutzer die Fähigkeit angeeignet hat, das Verhalten des Roboters vorherzusagen. Der Abfall der anthropomorphen Effekte kann jedoch durch unerwartetes Verhalten des Roboters immer wieder zeitweise unterbrochen werden. Nach dieser Gewöhnungszeit setzt die Stabilisierungsphase ein, in der die beobachtbaren anthropomorphen Effekte ein stabiles Niveau (SLA) annehmen.

Dieser dynamischen Perspektive fügen die Autoren nun noch ein ebenfalls drei Phasen durchlaufendes kognitives Modell der sukzessiven Bildung eines mentalen Modells der nicht-menschlichen Entität hinzu. Die erste Phase nennen sie instinktive, prä-kognitive Identifikation belebter Peers, welche zunächst durch die Beobachtung (v.a. der Form, Bewegung, Geräusche) des Roboters und Empathie[27] gekennzeichnet ist. Phase zwei wird durch die Projektion existierender (vertrauter) mentaler Modelle initiiert und entsteht durch kurze, nicht weiter kontextualisierte Interaktionen mit dem Roboter oder durch längere Beobachtungsepisoden (auch interaktiven Verhaltens), typischerweise in Laborumgebungen. Nach längerer kontextualisierter Interaktion mit dem Roboter (typischerweise zu Hause) ist der

26 Die mensch-zentrierten Faktoren lassen sich gut mit den dispositionalen, entwicklungsbezogenen und kulturellen Faktoren der Drei-Faktoren-Theorie von Epley et al. (2007) analogisieren.

27 Die Autoren zitieren Befunde der MRI-Forschung, welche auf die Relevanz der Spiegelneuronen, auch im Zusammenhang mit Robotern, verweisen.

Nutzer in der Lage, auf der Basis eigener Erfahrungen ein angepasstes mentales Modell zu kreieren, das nach wie vor anthropomorphe Züge aufweisen kann (Phase 3) (vgl. Lemaignan et al. 2014: 3f).

Mit diesem kognitiven Phasenmodell beschreiben die Autoren die grobe Abfolge dessen, was Epley et al. (2007) in ihrem Faktor „elicitation of agent knowledge“ als Zugänglichkeit anthropomorpher Repräsentationen, deren Aktivierung, Korrektur bzw. Integration und Anwendung fassen. Insgesamt ist ihr Modell gut kompatibel mit der Drei-Faktoren-Theorie, welche plausible Erklärungen für die verschiedenen Phasen anthropomorpher Effekte über die Interaktionsdauer hinweg liefern kann. So lässt sich das ICA etwa anhand dispositionaler, situationaler, entwicklungsbezogener und kultureller Faktoren fassen und der Neuigkeitseffekt (novelty effect) ließe sich in erster Linie durch die Effektanzmotivation bei Erstkontakt erklären. Für das SLA spielt hingegen die Sozialitätsmotivation womöglich die entscheidendste Rolle. Auch hier kann der Ansatz von Persson et al. (2000) die anthropomorphen Zuschreibungen inhaltlich differenzieren oder in irgendeiner Form quantifizieren, wie es für das „normalized level of anthropomorphic effects“ vorgesehen ist.

Im nächsten Kapitel werden die bisher vorgestellten Ansätze rekapituliert und integriert.

4 Integration

Es herrscht erstaunliche Homogenität hinsichtlich des Definitionskerns von Anthropomorphisierung, was vermutlich an der Etymologie und der Außeralltäglichkeit des Begriffes liegt. Beinahe alle Studien, die sich um eine Definition bemühen, referieren eine leichte Abwandlung einer Minimaldefinition, in der es um die Zuschreibung menschlicher Eigenschaften auf nicht-menschliche Entitäten geht. Manchmal wird Anthropomorphisierung als Zuschreibungstendenz beschrieben, andere Autoren legen Wert darauf, die Zuschreibung explizit als kognitiven Prozess zu fassen, wieder andere als Projektion von Erwartungen. Was zugeschrieben wird sind (meistens ganz allgemein formuliert) menschliche oder menschenähnliche Eigenschaften an nicht-menschliche (reale oder imaginierte) Agenten oder Objekte. In einigen Fällen werden diese eingegrenzt auf die physische Form oder das Verhalten; in anderen Fällen werden sie weiter spezifiziert als mentale Zustände (kognitiv oder emotional, wie Gedanken, Gefühle, Motivationen, Glauben) oder im Falle der Annäherung

über den umgekehrten Prozess der Dehumanisierung auch als spezifisch menschliche Eigenschaften wie Menschlichkeit. Wieder andere Definitionen fügen dem Zuschreibungsprozess noch eine Erklärung hinzu, indem sie darauf verweisen, dass es sich um eine Rationalisierung (Verstehen+Erklären) von beobachtbarem Verhalten handelt, indem eine vereinfachende Beschreibung, die sich jedoch jenseits bloßer Verhaltensbeschreibungen bewegt, angeboten wird.

Die besprochenen theoretischen Ansätze, die sich explizit mit Anthropomorphisierung beschäftigen, ergänzen sich inhaltlich sehr gut, indem sie unterschiedliche Aspekte des Phänomens beleuchten. Am umfassendsten betrachten Epley et al. (2007) den kognitiven Prozess, den sie als Inferenzprozess beim Beobachter konzeptualisieren. Qualitativ induktive Schlüsse nutzen bereits erworbenes Wissen und weiten es auf andere Gegenstandsbereiche aus. Im Falle von Anthropomorphisierungen wird also menschenbezogenes Wissen aktiviert und auf das Objekt der Anthropomorphisierung angewandt. Dieser Prozess kann durch die Effektanzmotivation (also die Motivation, effektiv mit der Umwelt zu interagieren) und die Sozialitätsmotivation (das Bedürfnis nach sozialer Zugehörigkeit) verstärkt werden. Die Autoren können mit ihrer Theorie Variabilitäten in den individuellen Anthropomorphisierungstendenzen erklären. Zudem formulieren sie dispositionale, situationale, entwicklungsbezogene und kulturelle Einflussmöglichkeiten auf die kognitiven und emotionalen Faktoren. Lemaignan et al. (2014) weisen hingegen auf die inhärente Dynamik des Anthropomorphisierungsprozesses hin, insbesondere im Falle der Langzeitinteraktion mit Robotersystemen. Diese Dynamik lässt sich anhand der Drei-Faktoren-Theorie erklären, da über die Interaktionsdauer hinweg neues bzw. detaillierteres Wissen über den Roboter erworben wird, welches menschenbezogenes durch roboterbezogenes Wissen ersetzen und die Anthropomorphisierungstendenz damit verändern kann. Außerdem ist ihr Hinweis auf den novelty effect von größter Bedeutung, da bislang die meisten HRI-Experimente im Laborkontext, also unter außeralltäglichen, höchst artifiziellen Bedingungen in Form eines Erstkontaktes stattfinden, wobei dieser Neuigkeitseffekt hinsichtlich der Anthropomorphisierung zum Zuge kommt. Die anfänglich hohen Anthropomorphisierungstendenzen können mit der Interaktionsdauer auf ein stabiles Niveau absinken. Unvorhersehbares, disruptives Verhalten des Roboters hingegen kann die Anthropomorphisierungstendenz immer wieder verstärken,

wenn es das adaptierte mentale Modell des Roboters in Frage stellt und mit menschenbezogenem Wissen anreichert (darauf wird später nochmals im Zusammenhang mit einem krisenexperimentellen Vorgehen in der MRI-Forschung eingegangen, vgl. Compagna/Marquardt 2015).

Was konkret zugeschrieben wird, spielt für einige Autoren eine untergeordnete Rolle. Epley et al. (2007) etwa gehen wie auch Dennett (1971) davon aus, mit ihrer Theorie sowohl schwächere als auch stärkere Formen der Anthropomorphisierung erklären zu können und betonen dabei die nicht zu unterschätzende verhaltensbezogene Wirkung anthropomorpher Zuschreibungen, auch wenn diese zunächst rein metaphorisch gebraucht zu werden scheinen. Dennoch liegt die Vermutung nahe, das all das, was unter der oben genannten Definition von Anthropomorphisierung in deren Gegenstandsbereich fällt, doch recht unterschiedliche Wirkungsgrade entfalten kann.

Was Anthropomorphisierung inhaltlich bedeutet bzw. wie sie qualitativ differenziert, beobachtet oder erfasst werden kann, bleibt in den besprochenen Ansätzen nach wie vor unterbeleuchtet. Den wichtigsten Hinweis geben Persson et al. (2000), indem sie darauf aufmerksam machen, dass Anthropomorphisierung auf verschiedenen Ebenen operiert, die über je spezifische Charakteristika verfügen und mit je spezifischen Erwartungsbündeln einhergehen. Interessant ist v.a. die Ebene der primitiven Kategorisierung, auf der bestimmte Cues spontane Anthropomorphisierungen auslösen können. Hier kann insbesondere auch anthropomorphes Design beim Roboter zum Einsatz kommen, denn menschenähnliche Gesichter oder Extremitäten, menschliche Bewegung(-sgeschwindigkeit) und insbesondere Stimmen können den Anthropomorphisierungsprozess in Gang bringen, indem sie menschenbezogenes Wissen aktivieren, das gemeinhin gut zugänglich ist. Diese Kategorisierung geht über die bloße Zuschreibung von Belebtheit (Animismus) hinaus und vollzieht sich zunächst unbewusst (Type I Processing bzw. automatisch-spontaner Modus der Informationsverarbeitung). Sie bildet die Basis für differenziertere und komplexere Zuschreibungen. Die folgenden Ebenen (primitive Psychologie, Alltagspsychologie, Traits bzw. Wesenszüge und soziale Rollen) implizieren im Prinzip Annahmen über die anthropomorphisierte Entität, die zunehmend komplexer werden und sich vom bloßen beobachtbaren Verhalten immer weiter entfernen.

Problematisch an diesem Modell ist, dass diese höheren Ebenen eigentlich nur auf verbale Äußerungen über die anthropomorphisierte Entität referieren können (bspw. post-hoc Beschreibungen/ Rationalisierungen der Interaktion; im Alltag: Erzählungen). Sie sind damit meist aus der Interaktion im Vollzug herausgelöst (außer der menschliche Interaktionspartner kommuniziert verbal mit der anthropomorphisierten Entität). Die anthropomorphe Zuschreibung muss in irgendeiner Form verbalisiert werden. In dieser Verbalisierung können dann Formulierungen enthalten sein, die Rückschlüsse auf unterschiedlich komplexe implizierte mentale Zustände zulassen (z.B. der Roboter will X würde als Aussage implizieren, dass der Roboter einen Willen hat und könnte darüber hinaus noch implizieren, dass man ihn für seine Handlungen verantwortlich machen kann, da er auf der Basis seines Willens Handlungsentscheidungen trifft).

An dieser Stelle liegt dann der Einwand nahe, dass diese Implikationen womöglich nicht unbedingt unmittelbar ihr handlungswirksames Potential entfalten, da es sich bei den entsprechenden Äußerungen doch eher um verbale Heuristiken handelt, die eine Interaktionserfahrung mit einem Robotersystem simplifizierend intersubjektiv kommunizierbar machen.[28] Das wäre genau das, was Dennett (1971) als den intentional stance bezeichnet, der das Verhalten eines als hinreichend komplex empfundenen Systems durch die Projektion bestimmter Geisteszustände erklärbar macht, ohne damit eine Aussage über den ontologischen Status der Entität vornehmen zu müssen. Dennoch sind auch diese Anthropomorphisierungen nicht uninteressant – sie können ihr handlungswirksames Potential auch sehr subtil entfalten, indem sie ein bestimmtes Framing der Situation begünstigen.[29]

Gerade im Falle von Robotern wird es allerdings interessant, wenn die Anthropomorphisierung nicht auf dieser Ebene verbleibt, sondern die Erwartungen, die aus den Zuschreibungen resultieren, in der Interaktion einen höheren Komplexitätsgrad annehmen. In diesem

28 Genau das beobachten Nass und Moon (2000) für Computer, aber auch Fussell et al. (2008) für Roboter. Sie berichten, dass die Versuchspersonen zwar soziales Verhalten gegenüber Computern zeigen bzw. bereitwillig anthropomorphisierendes Vokabular zur spontanen Beschreibung eines Roboters verwenden; sich bei expliziteren Nachfragen hinsichtlich der Implikationen dieses Verhaltens (also ob der PC als sozialer Akteur gelten kann bzw. ob der Roboter einen Geist, Gedanken, Motive etc. hat) jedoch dagegen wehren.

29 Waytz et al. (2010) etwa untersuchten die Implikationen von Anthropomorphismus hinsichtlich Moral, Verantwortung und normativ-sozialem Einfluss.

Falle übertritt die Interaktion die Schwelle zur sozialen Interaktion; dann fungieren Anthropomorphisierungen als interaktionsleitende Erwartungen, auf deren Basis sich Erwartungs-Erwartungen bilden können. Dann kann der Roboter als sozialer Akteur in einem genuin soziologischen Sinne begriffen werden. Die These lautet daher: Anthropomorphisierungen sind eine Vorstufe der Bildung von Erwartungs-Erwartungen. Das Spektrum anthropomorpher Zuschreibungen grenzt nach der hier vertretenen Auffassung am einen Ende an das, was als Animismus bezeichnet wird, also die Zuschreibung von Belebtheit auf Unbelebtes; auf der anderen Seite an was, was im soziologischen Sinne eine soziale Interaktion strukturiert: Erwartungs-Erwartungen. Damit erweist sich das Phänomen der Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung auch als soziologisch höchst interessant und anschlussfähig. Im letzten Abschnitt des Hauptteils werden daher nun nochmals verschiedene soziologische bzw. sozialtheoretische Perspektiven anklingen, welche ein besonderes Anschlusspotential an das Phänomen der Anthropomorphisierung bieten.

5 Soziologische bzw. sozialtheoretische Anschlüsse

Man kann sozialtheoretisch sehr basal ansetzen, um die Möglichkeit einer sozialen Interaktion mit Robotern zu plausibilisieren. George Herbert Mead etwa verweist auf die zentrale Bedeutung der Rollenübernahme für die Sozialisierung und Identitätsentwicklung beim Kind und weitet sie auch auf die physische Objektwelt aus (Mead 1934/2013). Die Rolle, die übernommen wird, kann diejenige eines signifikanten oder generalisierten Anderen[30] sein. Anthropomorphisierungen können nun in diesem Sinne als Prozesse der Rollenübernahme reformuliert werden. Mead bemerkt, dass auch unbelebte Gegenstände in die Haltung des verallgemeinerten Anderen einfließen können (Mead 1934/2013: 196). Ob die Perspektive eines Roboters auch als die eines signifikanten Anderen übernommen werden kann, ist jedoch eine Frage, die empirisch beantwortet werden müsste.

30 Charles Morris schreibt dazu im Vorwort zu Meads „Geist, Identität und Gesellschaft“: „der verallgemeinerte Andere ist jedweder andere, der als Einzelheit der Haltung der Rollenübernahme im jeweiligen kooperativen Prozeß gegenüber steht oder stehen könnte. Vom Standpunkt der Handlung aus gesehen, ist der verallgemeinerte Andere die Handlung der Rollenübernahme in ihrer Universalität.“ (Morris nach Mead 1934/2013: 31)

Rollen sind Bündel von Erwartungen. Anthropomorphisierungen als Zuschreibungen implizieren auch bestimmte Erwartungen hinsichtlich der zugeschriebenen Charakteristika (Zuschreibung und Erwartung sind im Prinzip zwei Seiten der gleichen Medaille). Als weiteren Anknüpfungspunkt eignet sich hierfür Gesa Lindemann, welche in ihrer Monographie „Das Soziale von seinen Grenzen her denken“ (Lindemann 2009) den gemeinsamen Nenner verschiedener sozialtheoretischer Ansätze auf den Begriff der Erwartungs-Erwartungen bringt. Sie nutzt den Erwartungsbegriff, da er soziologisch trennscharf, theorieneutral und gut operationalisierbar ist und entwirft in Anschluss an Plessner damit eine Theorie struktureller Komplexität. „Die konsensuelle sozialtheoretische Grundannahme besteht darin, die Komplexität der Beziehung zwischen den involvierten Entitäten konstitutiv für Sozialität aufzufassen.“ (Lindemann 2009: 164) Soziologisch interessant wird es, wenn Interaktionen einen Komplexitätsgrad aufweisen, der durch Erwartungs-Erwartungen auf beiden Seiten der interagierenden Entitäten strukturiert ist. Dabei überlässt die Theorie struktureller Komplexität es prinzipiell der Empirie, wer zum Kreis der sozialen Akteure zählt.

Aus einer konstruktivistischen Perspektive ließe sich nun argumentieren, dass es weniger darauf ankommt, ob beide interagierenden Entitäten diese Erwartungen tatsächlich formulieren, sondern vielmehr darauf, dass die handlungsleitende Orientierung einer Erwartungs-Erwartung (bzw. die handlungswirksame Durchsetzung einer Situationsdefinition) für die soziale Interaktion aufrechterhalten werden kann. Auch in der soziologischen Handlungstheorie gilt die Zuschreibungsperspektive mittlerweile als komplementär zum subjektiven Sinn (Schulz-Schaeffer 2009). Die Mensch-Roboter Interaktion kann aus der Zuschreibungsperspektive folglich als vereinseitigte soziale Interaktion gefasst werden. Die Interaktion unter der Bedingung räumlich-zeitlicher Kopräsenz, wie sie in MRI-Experimenten vorgesehen ist, eignet sich dabei in besonderem Maße, um das Problem des Fremdverstehens[31] anzuge-

31 In der phänomenologischen Lesart von Handlungen ist der Handelnde selbst letzte Instanz, die entscheiden kann, ob etwas eine vorentworfene Handlung war oder nicht. Gleichzeitig werden in der Lebenswelt des Alltags Entscheidungen darüber, ob jemand gehandelt hat, auf der Basis des gesellschaftlichen Wissensvorrats getroffen (Schulz-Schaeffer 2008: 211). Das wird ein großes Thema sein, wenn Roboter als handlungsfähig anerkannt werden sollten und in die Gesellschaft integriert werden. Verschiedene Ausprägungen des gesellschaftlichen Wissensvorrates im kulturellen Vergleich können hier zu sehr unterschiedlichen Lösungen führen, vgl. die Diskussion zum Autonomie-Sicherheits-Paradoxon bei Matzusaki und Lindemann (2015).

hen (vgl. Schulz-Schaeffer 2008: 214), worum es sich bei Anthropomorphisierungen als Rationalisierung beobachtbaren Verhaltens im Prinzip handelt.

Schütz und Luckmann formulieren die günstigen Bedingungen für Fremdverstehen in der Situation räumlich-zeitlicher Kopräsenz in den Strukturen der Lebenswelt folgendermaßen: „In der Begegnung ist mir das Bewußtseinsleben des anderen durch ein Maximum an Symptomfülle zugänglich. Da er mir leiblich gegenübersteht, kann ich die Vorgänge in seinem Bewußtsein nicht nur durch das, was er mir vorsätzlich mitteilt, erfassen, sondern auch noch durch Beobachtung und Auslegung seiner Bewegung, seines Gesichtsausdrucks, seiner Gesten, des Rhythmus und der Intonation seiner Rede usw.“ (Schütz/ Luckmann 1979: 95)

Anthropomorphe Zuschreibungen implizieren nun – mit variabler Stärke bzw. Differenziertheit – eben jene Bewusstseinszustände. Hier wird auch die herausragende Bedeutung der Verkörperung im Vergleich etwa zur Interaktion mit Softwareagenten deutlich. In der Interaktion mit einem Roboter, der bspw. durch einen sensorischen Apparat seine Umwelt wahrnehmen und durch entsprechende Programme differenziert darauf reagieren kann, werden in der natürlichen Einstellung der Alltagswelt Zuschreibungen möglich, die aus einer Projektion der Reziprozität der Perspektiven auf das Robotersystem resultieren. Dies kann durch anthropomorphes Design beim Roboter verstärkt werden, denn Augen bzw. augenähnliche Konstruktionen befördern bspw. die Erwartung, dass der Roboter seine Umwelt visuell wahrnimmt. Potenziert wird das durch die Nutzung des akustischen Kommunikationskanales (vgl. Meads (1934/2013) Bemerkungen zur vokalen Geste), insbesondere in Form von Sprache in Verbindung mit gestenbasierter Kommunikation.

Als letzter soziologischer Anschluss soll nun noch auf das erweiterte Modell der Frame Selektion (MdFS) von Hartmut Esser verwiesen werden (Esser 2006). Dieses lässt sich insbesondere mit der Frage nach der Handlungswirksamkeit von Zuschreibungen verknüpfen. Darüber hinaus kann es die kognitiven Prozesse und situationale Einflüsse gut modellieren. Als zentrale Stellschrauben geht es Esser um die Definition der Situation (Frame), welche die Handlungsentscheidung beeinflusst (1. Selektion) und den Modus der Informationsverarbeitung (2. Selektion: automatisch-spontan/AS-Modus oder reflektiert-kalkulierend/ RC-Modus). Im Falle der Anthropomorphisierung von Robotern geht es also um die Entscheidung: Schreibe ich einem Roboter menschliche bzw. menschenähnliche Charakteristika zu

(Frame i) oder nicht (Frame j)?

Beide Selektionen laufen keineswegs bewusst ab, sondern sind Modellierungen, welche die Funktion erfüllen sollen, Handeln zu erklären und zu prognostizieren. Essers Theorieauffassung nach sollen Theorien eben kein möglichst genaues Abbild der Realität sein, sondern in erster Linie ihren Zweck erfüllen – und das ist Erklärung und Prognose. Wenn sie das tun, sind sie sogar umso besser, je einfacher und sparsamer sie mit theoretischen Konzepten umgehen. Dieser Ansatz wird auch als Instrumentalismus bezeichnet, weil Theorien hier als mehr oder weniger brauchbare Instrumente zur Erklärung bestimmter Aspekte der Realität herangezogen werden – und nie der ganzen Realität (Esser 1999: 51f). Die Frame-Selektion wird als Wahl zwischen den beiden alternativen Frames i und j modelliert, welche als zentraler Größe vom *Match* abhängen, also der Passung der jeweiligen Situationsdefinition (vgl. Tabelle 2). Diese Passung setzt sich zusammen aus der Zugänglichkeit des Frames (also die Zugänglichkeit anthropomorpher Repräsentationen wie Epley et al. (2007) das beschreiben), dem Vorhandensein entsprechender signifikanter Symbole in der Situation (anthropomorphen *Cues*, bspw. durch anthropomorphes Design beim Roboter) und der Abwesenheit von Störungen. Der Match wird mit dem Erwartungsnutzen des jeweiligen Frames multipliziert, einem subjektiven Maß, das in diesem Falle den Vorzug einer anthropomorphen Zuschreibung in der konkreten Situation fasst (bspw. der Vorzug, der aus einer effektiven Auseinandersetzung mit der Umwelt resultiert, vgl. Effektanzmotiv).

	EU(i) = m U(i) EU(j) = (1-m) U(j)		EU(as) = s U(i) EU(rc) = p U(j) + (1-p) s U(i) – C
MATCH	m = a * e * u	REFLEXIONSBEDINGUNG	U(j) – s U(i) > $\frac{C}{p}$
RE-FRAMING	$\frac{U(j)}{U(i)} > \frac{m}{1-m}$		
SALIENZ	$s = \frac{m}{1-m} - \frac{U(j)}{U(i)}$		

Tabelle 2: Die beiden Selektionen im Modell der Frame Selektion (MdFS), eigene Darstellung nach Esser 2006, S. 149f

Legende:

- EU: expected utility (Erwartungsnutzen)
- m: Match (erwartete Passung der Situationsdefinition; a: Zugänglichkeit des Frames; e: Vorhandensein entsprechender signifikanter Symbole in der Situation; u: Abwesenheit von Störungen)
- as: automatisch-spontaner Modus der Informationsverarbeitung
- rc: rational-kalkulierender Modus der Informationsverarbeitung
- s: Salienz der Rahmung (Grad der Unempfindlichkeit der Modell-Selektion)
- p: Wahrscheinlichkeit durch rationales kalkulieren die „richtige“ Situationsdefinition herauszufinden
- U: subjektiver Nutzen/Bewertung des jeweiligen Frames
- C: Kosten der Reflexion

Ein weiteres Maß ist die Salienz, welche den Grad der Unempfindlichkeit der Frameselektion gegenüber Alternativen beschreibt. Sie fließt auch in die zweite Selektion, die Selektion des Modus der Informationsverarbeitung, mit ein. Hier wird modelliert, ob ein Prozessieren im AS- oder im RC-Modus angemessen ist. Reflektiertes Prozessieren ist immer mit gewis-

sen Kosten verbunden, die vom Gesamterwartungsnutzen abgezogen werden müssen. Abgesehen von diesen Kosten entscheidet sich die Wahl des reflektierten Modus durch die Nutzenwerte der jeweiligen Alternativframes, in Abhängigkeit von der Wahrscheinlichkeit, dass durch rationales Kalkulieren die „richtige“ Situationsdefinition gefunden wird. Die Reflexionsbedingung lässt sich damit umformulieren als Differenz der Bewertungen der jeweiligen Frames. Diese muss größer sein, als die Reflexionskosten im Verhältnis zur Erfolgswahrscheinlichkeit der Reflexion. (Da der Frame i der aktuelle Frame ist, wird dieser mit der Salienz multipliziert, die ja ein Maß für die Passung des aktuellen Frames ist.)

Die Ungleichung zeigt, dass Situationen, in denen die Reflexionskosten besonders hoch sind, wahrscheinlich im AS-Modus verarbeitet werden. Ebenso werden Situationen mit guter Passung, also großer Salienz, nicht in den RC-Modus übergehen.

Esser berücksichtigt nun auch Emotionen in seinem MdFS, indem er die Vermittlung zwischen Reizen, welche durch das sensorische System aufgenommen und verarbeitet werden, und der entsprechenden Reaktion durch das motorische System stark vereinfachend durch die Interaktion dreier Systeme modelliert (vgl. Abbildung 2). Das Verarbeitungssystem (intentionaler Pfad) ist für die Verarbeitung komplexer Situationen zuständig und bietet durch die serielle Informationsverarbeitung nur eine begrenzte Kapazität an. Hier entsteht Bewusstsein, hier werden Pläne und Intentionen gebildet und rationale Entscheidungen gefällt (RC-Modus). Die anderen beiden Systeme (kognitiver und emotionaler Pfad) gehören zum automatisch-spontanen Modus der Informationsverarbeitung. Hier finden sich verschiedene fertige Reaktionsprogramme, die Erregungen, Bewertungen und spontane Reaktionen auf bestimmte Stimuli auslösen. Im kognitiven System sind kognitive Repräsentationen für typische Situationen mit typischen Problemlösungen abgespeichert. Beide Systeme bieten schnell abrufbare Problemlösungen, die dafür aber dann stark standardisiert sind und nicht immer optimal passen. Alle Systeme werden gleichzeitig vom sensorischen System informiert und können unabhängig voneinander auf das motorische System Einfluss nehmen; die Verarbeitung dauert jedoch unterschiedlich lange. (Bsp.: Ein kindähnlich gestalteter Roboter löst vielleicht zunächst durch die starken visuellen Cues des Kindchenschemas eine direkte emotionale Reaktion aus, die dann mit gespeicherten Gedächtnisinhalten eine be-

stimmte, umsorgende Verhaltensweise befördert. Diese Reaktionen können bei der Aktivierung des Verarbeitungssystems jedoch als unangemessen empfunden und negiert werden.)

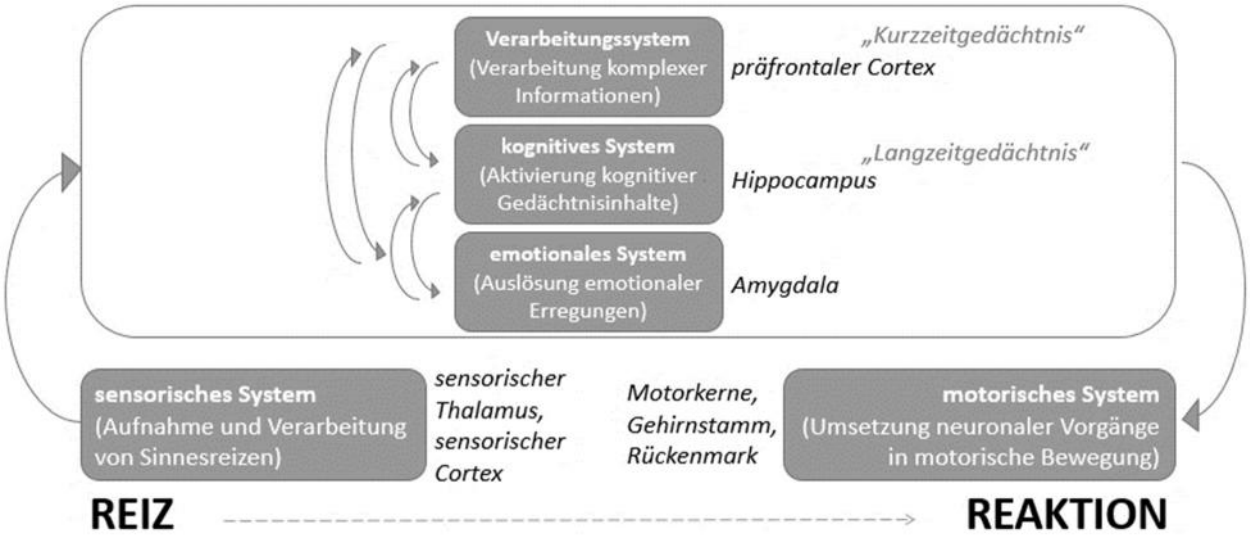


Abbildung 2: Reiz und Reaktion zwischen Emotion, Kognition und Bewusstsein, eigene Darstellung nach Esser 2006, S. 156

Essers Modell der Frame Selektion kann eine ganze Reihe von Befunden in der Anthropomorphisierungsforschung relativ gut erklären (bspw. die primitive Kategorisierung oder den Neuigkeitseffekt, aber auch die Bedeutung anthropomorphen Designs und individuelle Differenzen in den Anthropomorphisierungstendenzen). Es ist insbesondere auch mit der Drei-Faktoren-Theorie von Epley et al. (2007) vereinbar, wobei die Auslösung menschenbezogenen Wissens als kompetitiver Prozess zwischen dem kognitiven und dem Verarbeitungssystem modelliert werden kann; die motivationalen Komponenten fließen beim MdFS über die Nutzenterme der jeweiligen Frames mit ein. Der große Vorzug seines Modells ist das Konzept des (Mis-)Matches, welches konkrete situationale Cues mit gespeicherten mentalen Modellen abgleicht. Damit kann der hohen Relevanz des konkreten Interaktionskontextes Rechnung getragen werden. Gerade für experimentelle MRI-Forschung ist das von besonderer Relevanz, da entsprechende Instruktionen die Aktivierung bestimmter Gedächtnisinhalte wahrscheinlicher werden lassen (siehe Rahmenanalyse) und ein krisenexperimentelles Vorgehen, welches Erwartungen systematisch unterläuft, durch einen Mismatch einen Reflexionsprozess anstoßen kann, der damit die automatisch-spontane Reaktion besser beobachtbar und zugänglich macht (siehe Krisenexperimente) (Compagna/Marquardt

2015).

Esser modelliert jedoch nur die Handlungsentscheidung am Übergang der Logik der Situation zur Logik der Selektion. Lernprozesse, wie sie das dynamische Anthropomorphismusmodell von Lemaignan et al. (2014) betont, lassen sich damit schlecht fassen. Sie tauchen hier lediglich als veränderte kognitive Gedächtnisinhalte im Langzeitgedächtnis auf, welche bei Passung spontane Reaktionen befördern.

Ein integriertes, soziologisch anschlussfähiges Modell der Anthropomorphisierung sollte diese Dynamik unbedingt berücksichtigen. Darüber hinaus wäre es wünschenswert, die Stärke bzw. Differenziertheit der anthropomorphen Zuschreibung und ihre Handlungswirksamkeit miteinzubeziehen. Derart könnte die Anthropomorphisierung als Maß für den Grad an Sozialität einer Interaktion herangezogen werden.

Abbildung 3 zeigt ein Modell, welches die bislang besprochenen Ansätze zu integrieren versucht. Im Kern steht die Mensch-Roboter Interaktion als räumlich-zeitlich kopräsente, verkörperte Interaktion. Der Roboter wird vom menschlichen Interaktionspartner auf der Basis seiner Form, seines beobachtbaren Verhaltens und seiner Interaktionsmodalitäten interpretiert. Die Anthropomorphisierung als Zu-schreibung impliziert bestimmte Erwartungen hinsichtlich der zugeschriebenen Qualitäten. Durch das beobachtbare Verhalten des Roboters können diese Erwartungen bestätigt oder revidiert werden und auch die anthropomorphe Zuschreibung kann aufrechterhalten oder eingestellt werden. Darüber hinaus können situationale Cues eine Anthropomorphisierung befördern oder abschwächen. Der Zuschreib-ungsprozess wird beim menschlichen Interaktionspartner in erster Linie über das Wissen gesteuert. Im Falle einer anthropomorphen Zuschreibung wird menschenbezogenes Wissen aktiviert und auf den Roboter übertragen. Dieses Wissen kann je nach situationalen Cues unterschiedlich gut zugänglich sein. Wie die anthropomorphe Zuschreibung prozessiert wird, entscheidet sich anhand der Passung der situationalen Cues und der im subjektiven Wissensvorrat abgelegten mentalen Modelle und kann anhand des zuvor beschriebenen Modell der Frame Selektion modelliert werden.

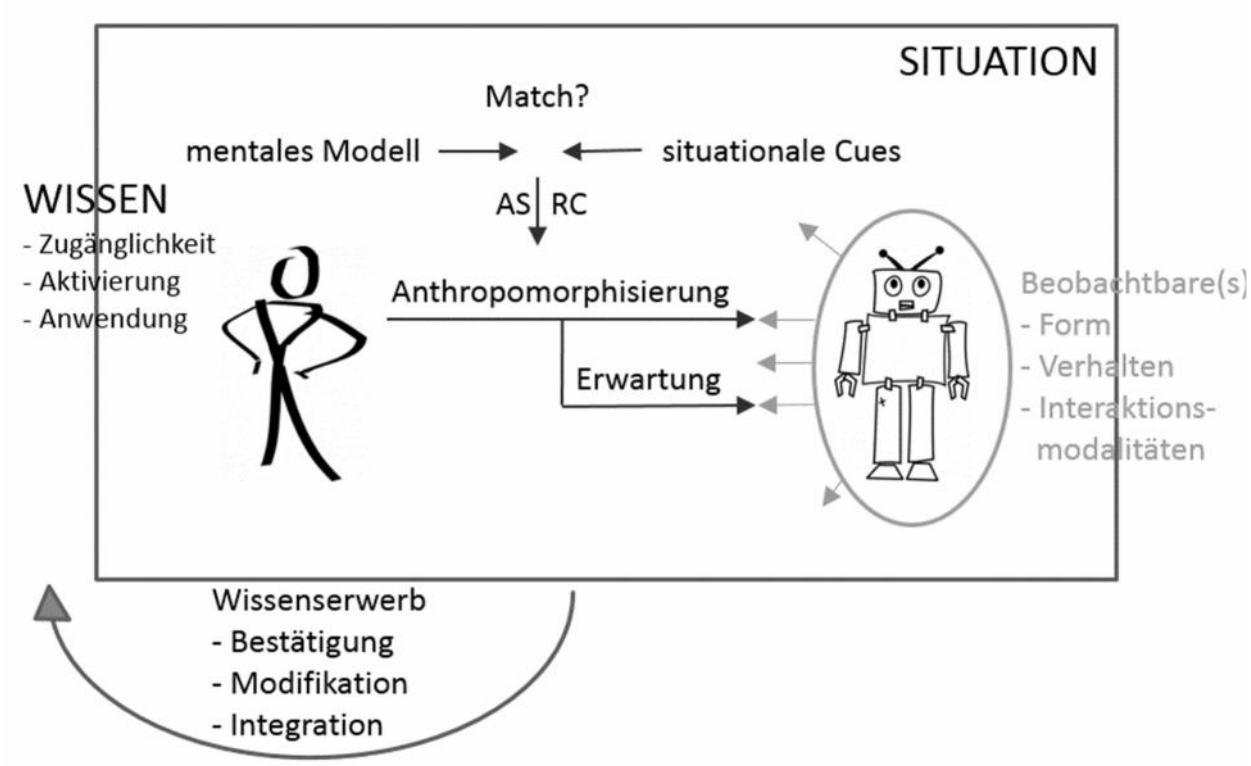


Abbildung 3: Modell der Anthropomorphisierung, Quelle: eigene Darstellung

Bei gutem Match erfolgt eine automatisch-spontane Reaktion. Ist der Match nicht ganz so gut, können zahlreiche andere Faktoren ins Gewicht fallen (bspw. das, was Epley et al. (2007) als „Need for Cognition“ bezeichnen; die Disposition, Spaß am reflektierten Denken zu haben, reduziert die Reflexionskosten; zudem können bei einem imperfekten Match dispositionale, entwicklungsbezogene oder kulturelle Einflüsse die Konstellationen der Nutzenterme für die Bewertung der Alternativframes „Anthropomorphisierung“ vs. „keine Anthropomorphisierung“ beeinflussen; insbesondere die Effektanzmotivation und die Sozialitätsmotivation verändern die subjektiven Wertungen einer anthropomorphen Zuschreibung). In der Interaktion wird durch die Bestätigung oder Enttäuschung von Erwartungen Wissen erworben, welches das mentale Modell des Roboters als menschenähnlich bestätigen, modifizieren oder korrigieren kann. Durch diese Rückkopplungsschleife über den Wissenserwerb erhält das Modell eine dynamische Perspektive. Zuletzt sei nun noch auf die Möglichkeit der Formulierung von Erwartungs-Erwartungen hinsichtlich des menschlichen Interaktionspartners verwiesen. Diese ist bedingt durch die inhaltliche Ausformulierung der anthropomorphen Zuschreibung, aus der die Erwartungen resultieren. Impliziert diese Zuschreibung,

dass der Roboter als Interaktionspartner ebenfalls Erwartungen an den menschlichen Interaktionspartner formuliert, so kann der Mensch hinsichtlich dieser zugeschriebenen Erwartungen des Roboters wiederum Erwartungen bilden, welche definitionsgemäß dann Erwartungs-Erwartungen wären.

Die Stärke der Anthropomorphisierung ließe sich nach inhaltlicher Differenziertheit (bspw. analog zu Lemaignan et al. 2014 oder Fussell et al. 2008, also welche Art menschlicher Eigenschaften sind impliziert? Eine alternative Kategorisierung könnte bspw. nach sensorischen, kognitiven, intentionalen, dispositionalen und sozialen Inferenzen differenzieren) und der Stärke des verhaltensbezogenen Einflusses, also der Handlungswirksamkeit der Zuschreibung vornehmen. Einen Hinweis auf die Stärke der Handlungswirksamkeit der Zuschreibung könnte eine im AS-Modus prozessierte (also spontane) Anthropomorphisierung sein, die auch im RC-Modus aufrechterhalten wird. Weitere Hinweise in MRI-Szenarien könnten verhaltensbezogene Maße, wie etwa die Befolgung von Bitten oder Aufforderungen des Roboters oder einstellungsbezogene Maße, wie die Zustimmung hinsichtlich verschiedener menschenähnlicher Charakteristika in Form einer Guttman-Skala oder eines semantischen Differentials sein. Wichtig bei fragebogenbasierten Methoden wäre es, den Befragten eine „trifft nicht zu“ Kategorie zur Auswahl zu stellen, damit die Anthropomorphisierung nicht als *forced choice* erfolgt.

6 Schluss

Anthropomorphisierung als Zuschreibung menschlicher Eigenschaften – im hier interessierenden Falle – auf Roboter wurde im Problemaufriss mit der Frage nach der Möglichkeit einer im soziologischen Sinne genuin sozialen Interaktion in Verbindung gebracht. Als doch recht alltägliches und niederschwelliges Phänomen lässt sich aus einer konstruktivistischen Perspektive hierfür die Frage nach der Handlungswirksamkeit anthropomorpher Zuschreibungen stellen.

Dieses Paper widmete sich anhand folgender Forschungsfragen dem Phänomen der Anthropomorphisierung:

1. Wie wird das Konzept der Anthropomorphisierung theoretisch gefasst?
2. Lassen sich ein gemeinsamer Kern rekonstruieren und verschiedene Perspektiven integrieren?
3. Welches Anschlusspotential bietet sich für eine soziologische bzw. sozialtheoretische Perspektive?

ad 1) Neben einigen ersten interessanten Befunden aus dem Forschungsüberblick, zeichnet sich der Großteil der Anthropomorphismusforschung doch eher durch einen unreflektierten und undifferenzierten Umgang mit dem Phänomen aus, was den Bedarf an theoretischen Schärfungen evident werden lässt. Daher wurden zunächst theoretische Ansätze rezipiert, die den Anthropomorphisierungsprozess mit verschiedenen inhaltlichen Schwerpunktsetzungen beschreiben.

Dennetts intentional stance beschreibt eine Grundhaltung, die zur Rationalisierung und Prognostizierung eines hinreichend komplexen Systems eingenommen werden kann, um dessen beobachtbares Verhalten anhand geistesbezogener Charakteristika erklärbar zu machen.

Epley et al. (2007) entwarfen eine Drei-Faktoren-Theorie des Anthropomorphismus, welche verschiedene dispositionale, situationale, entwicklungsbezogene und kulturelle Einflussvariablen auf die drei Faktoren Auslösung menschenbezogenen Wissens, Effektanzmotivation und Sozialitätsmotivation vorschlagen. Ihre Theorie fasst insbesondere den kognitiven Prozess des Wissenserwerbs, dessen Aktivierung und Anwendung auf das Objekt der Anthropomorphisierung, welcher durch motivationale Einflüsse gelenkt werden kann. Sie betonen die Bedeutung des Wissens und seiner situativen Zugänglichkeit. Zudem weisen sie auf schwächere und stärkere Anthropomorphisierungen hin, welche sie anhand der verhaltensbezogenen Konsequenzen charakterisieren.

Persson et al. (2000) leisten einen Beitrag zur inhaltlichen Differenzierung dessen, was bei Anthropomorphisierungen zugeschrieben wird. Sie konzeptualisieren Anthropomorphisierung als Phänomen, das auf verschiedenen Ebenen operiert. Diese reichen von der primitiven Kategorisierung (Bewegtheit, visuelle Erscheinung, Stimmen) über die primitive Psychologie (Selbsterhaltung, primäre Bedürfnisse, Triebe, Wahrnehmung), Alltagspsychologie

(Beziehung von Wahrnehmungen, Einstellungen, Zielen, Intentionen, Emotionen und Handlungen), Traits bzw. Wesenszüge (Generalisierung bestimmter stabiler Konstellationen innerer Zustände) bis zur Zuschreibung sozialer Rollen (normative Erwartungsbündel).

Lemaignan et al. (2014) fokussieren auf die inhärente Dynamik des Anthropomorphisierungsprozesses und modellieren anthropomorphe Zuschreibungen über die Interaktionsdauer hinweg anhand eines initialen Anthropomorphisierungskapitals (ICA), dem Neuigkeitseffekt (novelty effect) und eines stabilisierten Anthropomorphisierungslevels (SLA) über die drei Phasen der Initialisierung, Gewöhnung und Stabilisierung. Darüber hinaus weisen sie auf die Möglichkeit von zwischenzeitlichen Anthropomorphisierungsschüben hin, die durch unerwartetes Verhalten des Roboters entstehen können.

ad 2) Die Ansätze verfügen über einen gemeinsamen Definitionskern, in dem es um die Zuschreibung menschenähnlicher bzw. menschlicher Eigenschaften auf nicht-menschliche Entitäten geht und lassen sich daher gut integrieren. Während die Drei-Faktoren-Theorie die kognitiven Prozesse gut erklären kann, lässt sich mit Lemaignan et al. (2014) eine dynamische Perspektive auf das Phänomen Anthropomorphisierung einnehmen, das sich inhaltlich auf der Basis von Persson et al. (2000) weiter differenzieren lässt. Anthropomorphisierung kann als Phänomen zwischen Animismus und sozial wirksamen Zuschreibungen, aus denen Erwartungs-Erwartungen resultieren, angesiedelt werden.

ad 3) Als soziologische bzw. sozialtheoretische Anschlüsse wurde zunächst auf Mead (1934/2013) und das Konzept der Rollenübernahme verwiesen. Im Anschluss wurde mit Gesa Lindemanns Theorie struktureller Komplexität der gemeinsame Nenner verschiedener sozialtheoretischer Ansätze auf den Begriff der Erwartungs-Erwartungen gebracht (Lindemann 2009). Aus einer konstruktivistischen Zuschreibungsperspektive ist es weniger von Bedeutung, ob diese Erwartungen tatsächlich auf beiden Seiten der Interagierenden formuliert werden, sondern vielmehr, dass sie ihre handlungsorientierende Funktion (auf Seiten des menschlichen Interaktionspartners) entfalten. Zudem wurde auf die Bedeutung der räumlich-zeitlich kopräsenten, verkörperten Interaktion für das Problem des Fremdverstehens (Schütz/Luckmann 1979), für das Anthropomorphisierungen eine alltagsweltliche Lösung sind, verwiesen.

Abschließend wurde das erweiterte Modell der Frame Selektion von Hartmut Esser auf Anthropomorphisierungen angewandt. Anhand dieses Modells lassen sich über das Konzept des Matches die Relevanz der situationalen Cues und ihrer Passung zu gespeicherten mentalen Modellen verdeutlichen. Außerdem werden zwei Modi der Informationsverarbeitung aufgezeigt, in denen Anthropomorphisierungen prozessiert werden können: der automatisch-spontane/AS-Modus und der reflektiert-kalkulierende/RC-Modus der Informationsverarbeitung. Der sensorische Reiz und die motorische Reaktion werden durch die Aktivierung eines oder mehrerer Pfade (emotional, kognitiv, intentional) mediiert. Anhand dieses Modells lassen sich zahlreiche Befunde der Anthropomorphismus-forschung gut erklären. Auf dieser Basis wurde ein integriertes Anthropomorphismusmodell formuliert, das im Kern auf dem Modell der Frame Selektion (als Frage nach der Passung situationaler Cues und mentaler anthropomorpher Modelle) basiert und eine vereinseitigte (soziale) Interaktion zwischen einem Menschen und einem Roboter aus der Zuschreibungsperspektive modelliert. Im Kern stehen dabei aus Anthropomorphisierungen resultierende Erwartungen, die interaktiv stabilisiert oder enttäuscht werden können. Der Wissenserwerb aus den Interaktionserfahrungen fließt in die Stabilisierung oder Modifizierung des mentalen Modells (Frame) ein. Ist die Anthropomorphisierung inhaltlich derart differenziert, dass sie die Möglichkeit der Bildung von Erwartungen auf Seiten des Roboters impliziert, kann der menschliche Interaktionspartner Erwartungen hinsichtlich dieser Erwartungen formulieren, wobei es sich um Erwartungs-Erwartungen handeln würde. Die Stärke der anthropomorphen Zuschreibung kann bspw. durch eine inhaltliche Kategorisierung der Art der menschlichen Eigenschaften erfolgen; es kann auch die Stärke des verhaltensbezogenen Einflusses der Anthropomorphisierung (also deren Handlungswirksamkeit) herangezogen werden.

Zuguterletzt lässt sich an dieser Stellen nochmals auf das Potential einer fundierten Operationalisierung von Anthropomorph-isierung für die Evaluation der Mensch-Roboter Interaktion verweisen. Sollte es künftiger Forschung gelingen, ein adäquates Messinstrument für verschiedene Stärkegrade anthropomorpher Zuschreibungen (insbesondere auch im Zeitverlauf) zu entwickeln, könnten diese als Grad für die Sozialität der Interaktion herangezogen werden.

Literatur

Baecker, D., 2011: Who qualifies for communication? - A systems perspective on human and other possibly intelligent beings taking part in the next society. Technikfolgenabschätzung - Theorie und Praxis 20: 17–26.

Bartneck, C., D. Kulić, E. Croft und S. Zoghbi, 2009: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics 1: 71–81.

Bless, H., M. Wänke, G. Bohner, R.F. Fellhauer und N. Schwarz, 1994: Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben. Zeitschrift für Sozialpsychologie 25: 147–154.

Blythe, P.W., P.M. Todd und G.F. Miller, 1999: How motion reveals intention : Categorizing social interactions. S. 257–285 in: Simple heuristics that make us smart, New York: Oxford University Press.

Chin, M.G., V.K. Sims, B. Clark und G.R. Lopez, 2004: Measuring individual differences in anthropomorphism toward machines and animals. Bd. 48, S. 1252–1255 in: Proceedings of the human factors and ergonomics society annual meeting.

Chin, M.G., R.E. Yordon, B.R. Clark, T. Ballion, M.J. Dolezal, R. Shumaker und N. Finkelstein, 2005: Developing an anthropomorphic tendencies scale. Bd. 49, S. 1266–1268 in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Coeckelbergh, M., 2011: Humans, animals, and robots: A phenomenological approach to human-robot relations. International Journal of Social Robotics 3: 197–204.

Compagna, D. und M. Marquardt, 2015: Zur Evaluation von Mensch-Roboter Interaktionen (MRI) – ein methodischer Beitrag aus soziologischer Perspektive (Working Paper No. 03/2015). S. 4–18.

Dautenhahn, K., 2004: Socially intelligent agents in human primate culture. Agent Culture: Human-Agent Interaction in a Multicultural World., Lawrence Erlbaum Associates.

Dautenhahn, K., 2007: Methodology and themes of human-robot interaction: a growing research field. International Journal of Advanced Robotic Systems 4: 103–108.

Dennett, D., 2009: Intentional systems theory. S. 339–350 in: A. Beckermann, B.P. McLaughlin & S. Walter (Hg.), The Oxford handbook of philosophy of mind.

Dennett, D.C., 1971: Intentional systems. The Journal of Philosophy 68: 87–106.

Duffy, B.R., 2003: Anthropomorphism and the social robot. Robotics and autonomous systems 42: 177–190.

Echterhoff, G., G. Bohner und F. Siebler, 2006: “Social Robotics” und Mensch-Maschine-Interaktion. Zeitschrift für Sozialpsychologie 37: 219–231.

Epley, N., A. Waytz und J.T. Cacioppo, 2007: On seeing human: a three-factor theory of anthropomorphism. Psychological review 114: 864.

Esser, H., 1993: Soziologie: allgemeine Grundlagen., Frankfurt/Main ; New York: Campus.

Esser, H., 2006: Affektuelles Handeln: Emotionen und das Modell der Frame-Selektion. S. 143–174 in: R. Schützeichel (Hg.), Emotionen und Sozialtheorie. Disziplinäre Ansätze, Frankfurt am Main: Campus.

Fink, J., 2012: Anthropomorphism and human likeness in the design of robots and human-robot interaction. S. 199–208 in: International Conference on Social Robotics.

Fong, T., I. Nourbakhsh und K. Dautenhahn, 2003: A survey of socially interactive robots. Robotics and Autonomous Systems 42: 143–166.

Fussell, S.R., S. Kiesler, L.D. Setlock und V. Yew, 2008: How people anthropomorphize robots. S. 145–152 in: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction.

Goetz, J., S. Kiesler und A. Powers, 2003: Matching robot appearance and behavior to tasks to improve human-robot cooperation. S. 55–60 in: Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on.

Guido, G. und A.M. Peluso, 2015: Brand anthropomorphism: Conceptualization, measurement, and impact on brand personality and loyalty. Journal of Brand Management 22: 1–19.

Haslam, N., 2006: Dehumanization: An integrative review. Personality and social psychology review 10: 252–264.

Hegel, F., S. Gieselmann, A. Peters, P. Holthaus und B. Wrede, 2011: Towards a typology of meaningful signals and cues in social robotics. S. 72–78 in: 2011 RO-MAN.

Hegel, F., S. Krach, T. Kircher, B. Wrede und G. Sagerer, 2008: Understanding social robots: A user study on anthropomorphism. S. 574–579 in: RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication.

Heider, F. und M. Simmel, 1944: An experimental study of apparent behavior. The American Journal of Psychology 57: 243–259.

Hellen, K. und M. Sääksjärvi, 2013: Development of a scale measuring childlike anthropomorphism in products. Journal of Marketing Management 29: 141–157.

Kamide, H., F. Eyssel und T. Arai, 2013: Psychological Anthropomorphism of Robots. S. 199–208 in: G. Herrmann, M.J. Pearson, A. Lenz, P. Bremner, A. Spiers & U. Leonards (Hg.), Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings, Cham: Springer International Publishing.

Kiesler, S., A. Powers, S.R. Fussell und C. Torrey, 2008: Anthropomorphic interactions with a robot and robot-like agent. Social Cognition 26: 169.

Knorr-Cetina, K., 1998: Sozialität mit Objekten - Soziale Beziehungen in post-traditionalen Wissensgesellschaften. Bd. 42, S. 83–120 in: W. Rammert (Hg.), Technik und Sozialtheorie, Frankfurt a.M. [u.a.]: Campus.

Lemaignan, S., J. Fink, P. Dillenbourg und C. Braboszcz, 2014: The Cognitive Correlates of Anthropomorphism. in: 2014 Human-Robot Interaction Conference, Workshop„ HRI: a bridge between Robotics and Neuroscience“.

Lindemann, G., 1999: Doppelte Kontingenz und reflexive Anthropologie. Zeitschrift für Soziologie 28: 165–181.

Lindemann, G., 2009: Die Verkörperung des Sozialen - Theoriekonstruktion und empirische Forschungsperspektiven. S. 162–181 in: Dies. (Hg.), Das Soziale von seinen Grenzen her denken, Weilerswist: Velbrück Wissenschaft.

Matsuzaki, H. und G. Lindemann, 2015: The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. AI & Society1–17.

Mead, G.H., 1934/2013: Geist, Identität und Gesellschaft: aus der Sicht des Sozialbehaviorismus., Frankfurt am Main: Suhrkamp.

Meister, M., 2013: When is a Robot really Social? An Outline of the Robot Sociologicus. Science, Technology & Innovation Studies 10: .

Mithen, S.J., 1996: The prehistory of the mind: a search for the origins of art, religion, and science., London: Thames and Hudson.

Morewedge, C.K., J. Preston und D.M. Wegner, 2007: Timescale bias in the attribution of mind. Journal of personality and social psychology 93: 1.

Nass, C. und Y. Moon, 2000: Machines and mindlessness: Social responses to computers. Journal of social issues 56: 81–103.

Persson, P., J. Laaksolahti und P. Lönnqvist, 2000: Anthropomorphism–A multi-layered phenomenon. S. 131–135 in: Proc. Socially Intelligent Agents-the Human in the Loop, AAAI Fall Symposium, Technical Report FS-00-04.

Rammert, W. und I. Schulz-Schaeffer, 2002: Technik und Handeln - Wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt (Working Paper No. TUTS-WP-4-2002).

Reeves, B. und C. Nass, 1998: The media equation: how people treat computers, television, and new media like real people and places., Stanford, CA: CSLI Publ.

Reichertz, J., 2013: Die Abduktion in der qualitativen Sozialforschung: über die Entdeckung des Neuen., Wiesbaden: Springer VS.

Riek, L.D., T.-C. Rabinowitch, B. Chakrabarti und P. Robinson, 2009: How anthropomorphism affects empathy toward robots. S. 245–246 in: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction.

Ruijten, P.A., D.H. Bouten, D.C. Rouschop, J. Ham und C.J. Midden, 2014: Introducing a rasch-type anthropomorphism scale. S. 280–281 in: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction.

Salem, M., F. Eyssel, K. Rohlfing, S. Kopp und F. Joublin, 2013: To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. International Journal of Social Robotics 5: 313–323.

Schiffhauer, B., 2015: Determinanten von Anthropomorphismus und ihre Bedeutung für Dehumanisierung. Zuschreibung und Absprechen von Menschlichkeit gegenüber Menschen und nichtmenschlichen Entitäten, Universität Bielefeld.

Schmitz, M., 2011: Concepts for life-like interactive objects. S. 157–164 in: Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction.

Schulz-Schaeffer, I., 2008: Soziales Handeln, Fremdverstehen und Handlungszuschreibung. S. 211–221 in: J. Raab, M. Pfadenhauer, P. Stegmaier, J. Dreher & B. Schnettler (Hg.), Phänomenologie und Soziologie. Theoretische Positionen, aktuelle Problemfelder und empirische Umsetzungen, Wiesbaden: VS Verlag für Sozialwissenschaften.

Schulz-Schaeffer, I., 2009: Handlungszuschreibung und Situationsdefinition. Kölner Zeitschrift für Soziologie und Sozialpsychologie 61: 1–24.

Schütz, A. und T. Luckmann, 1979: Strukturen der Lebenswelt (Band 1)., Frankfurt am Main: Suhrkamp.

Semin, G.R. und K. Fiedler, 1991: The linguistic category model, its bases, applications and range. European review of social psychology 2: 1–30.

Serpell, J.A., 2002: Anthropomorphism and Anthropomorphic Selection—Beyond the„ Cute Response“. Society & Animals 10: 437–454.

Turkle, S. (Hg.), 2007: Evocative objects: things we think with., Cambridge, Mass: MIT Press.

Waytz, A., J. Cacioppo und N. Epley, 2010: Who sees human? The stability and importance of individual differences in anthropomorphism. Perspectives on Psychological Science 5: 219–232.

Weiss, A. und C. Bartneck, 2015: Meta analysis of the usage of the Godspeed Questionnaire Series. S. 381–388 in: Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on.

Young, J.E., J. Sung, A. Voida, E. Sharlin, T. Igarashi, H.I. Christensen und R.E. Grinter, 2011: Evaluating human-robot interaction. International Journal of Social Robotics 3: 53–67.

Złotowski, J., 2015: Understanding Anthropomorphism in the Interaction Between Users and Robots.